MULTI-DIMENSIONAL CHROMATOGRAPHY SYSTEM FOR PROTEOMICS

The present invention relates generally to methods and compositions for identifying
5   biomolecules present in a biological sample. More specifically, the present invention relates to
mass spectrometry-based methods and compositions for identifying proteins present in a body
fluid.

BACKGROUND OF THE INVENTION

The analysis of a proteome involves the separation of the proteins in a sample followed
10 by the identification of the resolved proteins, a challenging task given the tremendous chemical
heterogeneity in virtually all parameters that can be measured. For example, some cytokines
weigh between 1 and 2kD, while large muscle proteins weigh close to 1000kD. Some proteins
are very soluble and may also be present at high concentration in aqueous media (e.g. albumin,
40mg/mL in plasma), while some membrane proteins have more than 75% of their amino acids
15 buried in the phospholipid bilayer and are thus not amenable to studies in aqueous solvents.

Moreover, proteins are present in extremely divergent concentrations in cells, ranging
from 100 to 100,000,000 molecules per cell. The large range of concentration presents
considerable difficulties for protein analysis since available methods tend to lack the resolution to
detect non-abundant proteins – that is, current methods are often overloaded with protein if
20 practiced with the large sample volumes that would be required in order to detect low abundance
proteins and peptides. The term "dynamic range" is used in the art in the context of biological
samples, referring to the ratio of the concentrations of the most abundant molecules to the least
abundant ones. For example, human blood proteins can differ in their concentrations by a factor
of 10^9 (10^9 pg/ml for serum albumin compared with 0-5 pg/ml for interleukin 6).

25   Analysis of a proteome may allow the identification of proteins useful as protein
therapeutics, as biological targets for intervention via an interacting molecule, or as biomarkers
for the characterization of tissue and diagnosis of disease. Biochemical markers, for example,
can be identified by analyzing tissue or body samples from a subject, preferably a mammal, with
the disease of interest and then comparing the results of the analysis with those obtained from a
30 subject without the disease. One successful approach using two-dimensional gel electrophoresis
has led to the identification of a variety of marker proteins that are present at a significantly
different concentration in tissue or body fluid samples of a diseased mammal relative to a normal

mammal. See, for example, Partin et al. (1993) Cancer Res. 53:744-746 which describes the identification of prostate cancer markers and Getzenberg et al. (1996) Cancer Res. 5 6:1690-1694, which describes the identification of bladder cancer markers.

Two-dimensional polyacrylamide-gel electrophoresis (2D PAGE) followed by mass
5   spectrometry (MS) is the most widely used method of protein resolution and identification. In 2D PAGE, proteins are separated in one dimension by isoelectric point (pI) and in the other dimension by apparent molecular weight. As a result, a single 2D-PAGE system can resolve more than 1500 proteins. However, 2D-PAGE has significant disadvantages. Firstly, each spot from 2D-PAGE must be individually extracted, digested and analyzed which is cumbersome,
10   although limited robotic means have recently been developed to help automate this process. 2D-PAGE also has a limited loading capacity and limited detection limit for staining, which, due to the large dynamic range of protein abundance in biological samples, results in an inability to analyze a complete proteome. Two-dimensional gel electrophoresis and tandem mass spectrometry typically have a dynamic range of only 10^2 to 10^4.

15   As an alternative, methods not based on gel electrophoresis have been studied. It has generally been thought that a substitute to 2D-PAGE should resolve proteins as well as 2D-PAGE and also allow the rapid identification of resolved proteins. Attempts directed at alternative methods have included one and two-dimensional (1D and 2D) chromatography methods using high performance liquid chromatography (HPLC), capillary isoelectric focusing (CIEF), capillary
20   electrophoresis (CE) or micro capillary chromatography. These methods would thus eliminate cumbersome transfer steps from the 2D-PAGE device to the mass spectrometer. While chromatography devices and method have been in use for many years, most existing systems have been developed for the purification of a particular protein of interest rather than for separating large numbers of proteins. Since separation conditions vary for each protein, the
25   chromatography based systems configured to date do not teach how to treat a sample in such a way that large numbers of different proteins, particularly low-abundance proteins, are obtained in sufficient quantity and with sufficient purity so that they can be characterized by mass spectrometry.

In the simplest liquid chromatography based methods, non-PAGE means have included
30   1D chromatography and MS. In one example, CIEF and electrospray ionization (ESI) MS have been investigated (Tang et al, (1997) Anal. Chem. 69: 3177-3182). However Tang et al resolved very few peptides electrophoretically. Two-dimensional chromatographic methods linked to MS

or MS-MS have been developed as well. Several examples of systems for protein identification using peptide mapping have been described. In one example, Raida et al ((1999) J. Am. Soc. Mass Spectr. 10: 45-54), analyzing peptides from human plasma filtrate by cation-exchange chromatography followed by reverse-phase HPLC, detected over 3000 distinct peptide masses but

5   were able to determine the identity of only relatively few peptides. Wall et al. ((2000) Anal. Chem. 72: 1099-1111) examined a human erythroleukemia cell lysate using IEF apparatus and separating fractions obtained thereby using reverse phase HPLC column. Again, Wall et al positively identified only very few (38) proteins. Opiteck et al provide a system using a cation exchange column eluted in stepwise fashion onto a reverse phase column (Opiteck et al, (1997)

10   Anal. Chem 69: 1518-1424). In another system, Opiteck et al coupled size exclusion and reverse phase HPLC (Opiteck et al, (1998) Anal. Biochem. 258: 349-361). However, in both methods of Opiteck et al, only a low number of proteins were identified from E. coli lysates.

Other methods have focused on coupling chromatographic means to MS-MS. In one system, E. coli proteins were fractionated by anion exchange HPLC and portions were digested

15   with trypsin and processed on a reverse phase microcolumn HPLC (Link et al, (1997) Int. J. Mass. Spectrom. Ion Proc. 160: 303-316). In another example, a mixture from Saccharomyces cerevisiae ribosomes was loaded onto reverse phase and eluted onto a cation exchange column from which peptides were separated and sprayed into an ESI tandem mass spectrometer (Tong et al, (1999) Anal. Chem. 71: 2270-2278). In a further system, a peptide mixture from

20   Saccharomyces cerevisiae was loaded onto a biphasic 2D column packed with cation exchange and reverse phase materials and eluted onto an ESI MS-MS (Link et al. (1999) Nat. Biotechnol. 17: 676-682). Nevertheless, these methods have not allowed a substantially complete detection of proteins in complex mixtures from biological samples. One presumed factor in the failure of previous methods is that the known liquid chromatography systems have not allowed for the

25   analysis of intact (e.g. undigested) proteins, thereby rendering protein identification considerably more difficult. Moreover, none of the systems have been shown to have adequate resolution to detect low abundance proteins in a mixture having a wide dynamic range of protein concentrations such as human body fluids. Thus, as is the case with 2D-PAGE methods, to date no liquid chromatography-based system has been described providing a means to analyze a

30   complete proteome from a sample as complex as a body fluid. As a result, the proteins identified in these studies are either relatively abundant proteins, or very few (e.g. highly incomplete analysis of a proteome), or proteins from less complex samples such as cell lysates.

There is, therefore, still a need in the art to develop new methods and systems that can be used to rapidly identify the biomolecules present in a complex biological sample. There is thus a need to complement existing methods for analyzing a complete proteome, identifying disease markers and low abundance proteins present in actual tissue or body fluid samples.

5

SUMMARY OF THE INVENTION

Systematic investigations to characterize human plasma proteins have been carried out for a considerable amount of time (Hochstrasser, D.F. (1993)). However, no system has been described which would be capable of analyzing the biomolecules or preferably proteins present in
10   a highly complete sample of biological fluids, particularly plasma. Some systems to date have used biological samples that do not include the entire range of proteins and peptide naturally present in a sample, such as plasma subjected to ultrafiltration (PCT patent publication WO 98/07036). This technique, however, tends to remove many proteins of interest, both of higher and lower molecular weight. Other systems have been available only for characterization of
15   proteins present in cell lysate samples, and moreover would not have the sensitivity to detect proteins present at low concentrations.
The present inventors have achieved a separation of the biomolecules in a biological sample using novel means allowing the analysis of proteins present at low concentrations. Endocrine factors (e.g., growth factors and hormones) and active peptides (e.g., proteolytic products and post-
20   translationally modified protein species) are usually present at very low concentrations and are difficult to characterize genetically. These polypeptides clearly play a role in, and thus are valuable indicators of, the health of an individual. Furthermore, these polypeptides are generally small and easy to synthesize. The invention provides methods for complete fractionation of a biological sample, as opposed to traditional methods focused on purification of a particular
25   protein or set of proteins. The methods involve separating proteins from a fluid sample by size exclusion chromatography into a limited number of fractions, preferably 10, more preferably 5, even more preferably 2 fractions, and then further separating small proteins using other chromatographic means. The chromatographic separation protocol is adapted for the separation or fractionation of protein fractions containing substantially intact (e.g. undigested) proteins and
30 · peptides, devoid of large proteins, and having reduced content of one or more high-abundance proteins. This provides a sensitive and comprehensive means for detecting low abundance proteins. Larger proteins are separated and analyzed in parallel with the help of electrophoretic means if desired.

The methods of the invention, unlike currently available methods such as 2D PAGE, are scalable. Thus, a wide range of different sample volume can be used, including large (e.g. at least several liters) quantities of biological samples such as plasma.

The invention features methods of separating, identifying and/or analyzing proteins of a sample wherein a size exclusion column is used to select small proteins, *i.e.* proteins with a molecular weight below a preset cutoff of, e.g., 30kD, 20kD, or 15kD. The methods can be exemplified in a number of embodiments.

One embodiment is a method of separation, detection or analysis of proteins of a sample which includes:

- introducing a sample to a size exclusion column comprising an input and an output end, wherein the size exclusion column separates proteins of the sample and retaining the proteins of less than 30kD, less than 15kD, most preferably less than 20kD in a limited number of output fractions (preferably 10, more preferably 5, still more preferably 2, even more preferably 1); and

- introducing said retained output fractions from said size exclusion column to at least one further chromatographic column which separates proteins according to one or more additional physical property(ies) wherein said at least one further chromatographic column produces an effluent stream of eluted components (first or more dimension(s) of separation).

In preferred methods of the invention, improved sensitivity is achieved when using the system of the invention by incorporating a step to remove one or more of the most abundant proteins from the biological sample. Reducing the content of a protein in a sample can also be referred to as 'depleting' the sample of a protein. Preferably as much of the abundant protein is removed as possible, although the term 'depletion' does not require complete removal. Preferably, a biological sample comprising a target protein is introduced to a functionalized support, preferably an affinity column, including input and output ends and a target-specific adsorbing means, wherein the functionalized support selectively retains the target protein from the sample to produce an effluent stream having reduced concentration, or substantially lacking, the target protein. Thus, in a preferred aspect, the method comprises:

- introducing a sample to one or more affinity chromatography column(s), each of them functionalized with a ligand capable of selectively or specifically binding one abundant protein in the sample, wherein the affinity column(s) produce(s) or output(s) a sample having a reduced content in at least one abundant protein;

- introducing said sample having a reduced content in at least one abundant protein to a size exclusion column comprising an input and an output end, wherein the size exclusion column separates proteins of the depleted sample and retaining the proteins of less than 30kD, less than

15kD, most preferably less than 20kD in a limited number of output fractions (preferably 10, more preferably 5, still more preferably 2, even more preferably 1); and

- introducing said retained output fractions from said size exclusion column to at least one further chromatographic column which separates proteins according to one or more additional

5   physical property(ies), wherein said at least one further chromatographic column produces an effluent stream of eluted components (first or more dimension(s) of separation).

In the methods of the invention, preferred examples for said further column(s) (e.g. additional dimension(s)) include passage over ion exchange and/or reverse phase HPLC columns, most preferably an ion (cation) exchange followed by reverse phase HPLC. For example, where

10   two further columns are used, the first further column separates components to produce a first effluent stream which is split in a number of fractions (first dimension of separation), and then the fractions are individually introduced to the second further column to produce a second effluent stream (second dimension of separation).

Thus, in a further preferred aspect, the method comprises:

15   - introducing a sample to one or more affinity chromatography column(s), each of them functionalized with a ligand capable of selectively or specifically binding one abundant protein in the sample, wherein the affinity column(s) produce(s) or output(s) a sample having a reduced content of at least one abundant protein;

- introducing said sample having a reduced content of at least one abundant protein to a

20   size exclusion column comprising an input and an output end, wherein the size exclusion column separates proteins of the depleted sample and retaining the proteins of less than 30kD, less than 15kD, most preferably less than 20kD in a limited number of output fractions (preferably 10, more preferably 5, still more preferably 2, even more preferably 1);

- introducing said retained output fractions from said size exclusion column to an ion

25   exchange chromatography column, wherein the ion exchange column produces a first effluent stream of eluted components; and

- introducing eluted components from said first effluent stream to a reverse phase HPLC column, wherein the reverse phase HPLC column produces a second effluent stream of eluted components.

30   In preferred embodiments of the methods, all the output fractions from said size exclusion column which comprise, or are enriched in, proteins of less than 30kD, less than 15kD, most preferably less than 20kD, are introduced to a further chromatography column. Treating the complete set of proteins under the molecular weight cutoff provides a more complete inventory of low molecular weight proteins from a proteome.

In further preferred aspects of the methods, the components introduced to said affinity column, said size exclusion column, said further column(s), said ion exchange column and/or said reverse phase column are substantially intact proteins. That is, said proteins have not been subjected to a digestion step. Most preferably, in addition, the biological sample has been

5   maintained so as to prevent unwanted degradation or digestion of proteins.

Furthermore, in any of the methods of the invention, the method comprises concentrating the components eluted from the size exclusion column. In preferred aspect, the methods thus include introducing retained fractions from said size exclusion column, which comprise, or are enriched in, proteins of less than 30kD, less than 15kD, most preferably less than 20kD, to a

10   means for increasing the concentration of proteins in solution. Preferably said means is a reverse phase capture column.

In further aspects of any of the methods of the invention, the methods further comprise:

-       directly detecting the mass of a plurality of the proteins of less than 20kD eluted from one of said chromatographic column, preferably using mass spectrometry.

15   In a preferred embodiment, the methods of the invention are carried out on a biological sample from individuals displaying a detectable trait, for example a disease of interest, and repeated for a biological sample obtained from individuals not displaying a detectable trait, to be used as a reference sample.

The invention thus features a system and methods for the rapid and efficient separation of

20   proteins and other biological macromolecules. In one aspect, the invention provides a means for total analysis of a protein-containing sample comprising a means for specifically removing abundant proteins, a means for separating lower and higher molecular weight proteins, and at least one chromatography column for fractionation of lower molecular weight proteins and peptides. In preferred aspects, the system also comprises a gel electrophoresis means for

25   separation and recovery of higher molecular weight proteins. Preferably the system further comprises a mass spectrometer or an array of mass spectrometers and an analysis means for identification of peptides and proteins from mass spectrometric data.

In one aspect, provided is a system including a sample input means, a size exclusion separation device for separating proteins in a limited number of outputs fractions (preferably 10,

30   more preferably 5, even more preferably 2); at least one liquid chromatography column for separating proteins according to an additional physical property. Preferably the column for separating proteins according to an additional physical property is operative successively with the size exclusion separation device. Most preferably the system further provides a means for the selective, or more preferably specific, removal of one or more abundant proteins before the size

7

exclusion separation. Most preferably the system further comprises at least a size exclusion separation device and at least two further chromatography columns adapted to separate proteins according to a distinct dimension. Most preferably said further columns comprise an ion exchange column and a reverse phase HPLC column. Even more preferably said further columns

5 comprise an ion exchange column and two reverse phase HPLC columns, in order to separate the proteins according to 3 dimensions. The various means and steps for separation of proteins in a sample and/or elimination of abundant proteins can be arranged in any suitable order. However, preferred arrangements are further disclosed below.

According to preferred embodiments as further described herein, the preferred system

10 comprises: a sample input means; one or more solid support, preferably comprised in an affinity chromatography column, said support functionalized with a ligand capable of binding an abundant protein in biological fluids; a size exclusion chromatography column; and at least one further liquid chromatography column. The system typically also comprises an injection valve connecting the sample input means to the first column of the system, and one or more pump

15 means for providing pressure delivery of a solution to the column(s). Preferably the pressure is provided via a valve (preferably a multiport valve). The system preferably also comprises a program means for specifying system control programs. In preferred embodiments, the apparatus further includes control means in communication with the pump means for controlling the pressure of delivery of the solution.

20 One or more of the columns can be operably linked, such that eluant from one column can be directly passed to a subsequent column, thereby also allowing at least a portion of the steps to be automated or carried out 'inline'. In preferred aspects, operably linked columns comprise a valve for removing at least a portion of the eluate, for example, for the purpose of storage.

The system may also comprise solution input means including plural solution reservoirs,

25 and a mixing valve, connecting the solution input means to a sample input means connected to a column, operative to mix solution from the reservoir, wherein program means specifies the mixing of solution by the mixing valve, and the delivery of the mixed solution to the column via a multiport injection valve.

The system generally also includes at least one detector means for detecting and

30 recording column output. When multiple columns are used according to preferred methods described herein, it may be preferable to employ a number of detector means to monitor output from a plurality of columns.

In another embodiment, the invention features a system for the separation of proteins, which includes a plurality of chromatography columns according to the invention, means for

introducing a solution into a first said column, multiple multi-port valves in communication with said columns through which solution is transported, and output means comprising a detector and data collector.

Preferably, this embodiment includes pump means for introducing solution, preferably a complex biological fluid such as plasma, into the first column; control means in communication with the pump means for controlling the pressure of delivery of the solvent. It will be appreciated that fractions or samples eluting from one column or support can be further processed inline or can be manually introduced to a further column. It will also be appreciated that fractions or samples can be aliquoted and/or pooled (e.g., from multiple, repetitive runs) and/or stored after eluting from any column, e.g. for use in another protocol, such as 2-dimensional gel electrophoresis. This would, for example, allow for comprehensive detection of proteins in a biological sample involving distinct protocols adapted for the separation and detection of lower and higher molecular weight proteins, and could involve in that case, in addition to the methods described above, introducing a sample having a reduced content of at least one abundant protein to a two-dimensional electrophoresis separation means and detecting the mass of a plurality of proteins separated using said two-dimensional electrophoresis means, preferably using mass spectrometry.

Preferably, at least two of the columns may be operably linked. The multiple valves include first, second, and third valves, and the solution is introduced through the first valve into the first column, through the second valve into the second column, and through the third valve to a third column, or to an output means. The valves are preferably multi-port valves. The solution introducing means comprises sample input means which includes a sample reservoir. The sample input means may further include a sample pump, and the solution introducing means may include plural solution reservoirs, a valve for selection and mixing solutions, and a pump for delivering solution to the first column. The output means may further include a fourth multi-port valve connecting the detector to the data collector. The detector may be a UV detector. The output means may further include a pH/conductivity detector in communication with the detector and the data collector through the fourth multi-port valve.

Preferably, the column has a first and a second end and at least one of the first or second columns is packed with a chromatography matrix which confers on the packed column a desired transit time and a separation according to a desired physical property of proteins.

In preferred embodiments, the system of the invention (e.g. the plurality of columns and supports and the arrangement thereof) is adapted for the separation of undigested lower molecular weight proteins, preferably under 30kD, more preferably under 15kD or most preferably under

20kD. In preferred aspects, a so-adapted system comprises a column functionalized with ligands specific for at least one abundant protein (for example an abundant plasma protein), a size exclusion means to select for lower molecular weight proteins, and at least one or preferably at least two further chromatography columns, each of said further columns packed with a particulate

5   matrix separation medium for separation according to a distinct protocol with the use of according elution media and conditions. Preferably the system also comprises a means for concentrating samples eluting from the size exclusion means; in a preferred aspect a reverse phase capture column is used. The system also includes a sample input means including an input valve for delivery of solutions to one of said further columns at the column input end and a

10   multiport valve for mixing solutions provided to the input valve, column output means for detecting column output including means for detection and providing a signal indicative thereof, and control means for operating the sample input means.

The control means may include a means for alternatively utilizing one of the chromatography columns or another, for example, or for delivering different media to different

15   columns, or for directing the processing of fractions obtained from a first fractionation protocol. The apparatus may thus include program means for specifying a sequence of separation process control programs to be successively run during operation. The program means may specify a separation program in which a first and a second column are utilized successively for separating proteins in the sample, or a separating program where multiple fractions of a sample are delivered

20   to a column (e.g. multiple runs), or a pooling program where a plurality of fractions are to be combined in a pooled fraction.

Preferably, one of the said further columns of the system includes an ion exchange chromatography matrix, most preferably a cation exchange matrix. Alternatively or additionally, one of the said further columns of the system column may include a reverse phase

25   chromatography matrix, preferably for use in reverse phase HPLC. Each of the further columns, (e.g. at least a first and a second further column) individually, may be removable and replaceable. The system of the invention may involve arranging columns so that they are operably connected. Nevertheless, in several preferred embodiments of the invention, samples and fractions will need to be processed according to more than one method or specific protocol. For example, at some

30   stages as described further below, it will be preferable to remove a portion of a sample or fractions for storage. At other stages, a separation step may have output fractions of proteins that are optimally further processed according to differing protocols. In these embodiments, columns may not be operably connected, or alternatively, more complex connections or manipulations are provided with the aid of multiport valves and a process control program.

In one embodiment of the invention, a monitoring device can be provided. At any stage of the present process, a sample can be arranged to flow through a detector, thereby resulting in a graph with peaks for proteins and other non-target solutes, from which a measure of the purity and concentration of the proteins can be derived. Advantages of this embodiment of the

5   invention include rapid monitoring of the presence, quantity, and/or purity of proteins in a fraction derived from a complex biological sample such as plasma. Impurities may also be monitored – such as certain abundant proteins, nucleic acids, endotoxins, or generally any biological molecule detectable in the sample.

The methods and apparatus of the invention are rapid, reliable, and adaptable, and are

10   particularly useful in the separation and purification of proteins, particularly in the context of identifying a large number of proteins from a biological sample by fluid separation. Further details and examples are provided below.

DETAILED DESCRIPTION OF THE DRAWINGS

15   Figure 1 is a flow diagram depicting one implementation of the methods of the invention. More precisely, steps 1 to 5 of Example 1 are illustrated.

Figure 2 is a flow diagram depicting one implementation of the methods of the invention. More precisely, steps 5 to 6 of Example 1 are illustrated.

Figure 3 illustrates the MALDI-MS spectra obtained from the protein Leptin using the

20   methods of the invention, as described in Example 2.

Figure 4 illustrates the MS/MS spectra obtained from the protein Leptin using the methods of the invention, as described in Example 2.

Figure 5 illustrates the sequence coverage obtained from the protein Leptin by mass spectrometry using the methods of the invention, as described in Example 2.

25

DETAILED DESCRIPTION

Definitions

As used herein, the term "polypeptide" refers to a polymer of amino acids without regard to the length of the polymer; thus, peptides, oligopeptides, and proteins are included within the

30   definition of polypeptide. This term also does not specify or exclude post-translational modifications of polypeptides, for example, polypeptides which include the covalent attachment of glycosyl, acetyl, phosphate, amide, lipid, carboxyl, acyl, or carbohydrate groups are expressly encompassed by the term polypeptide. Also included within the definition are polypeptides which contain one or more analogs of an amino acid (including, for example, non-naturally

occurring amino acids, amino acids which only occur naturally in an unrelated biological system, modified amino acids from mammalian systems etc.), polypeptides with substituted linkages, as well as other modifications known in the art, both naturally occurring and non-naturally occurring.

5   The term "protein" as used herein may be used synonymously with the term "polypeptide" or may refer to, in addition, a complex of two or more polypeptides which may be linked by bonds other than peptide bonds, for example, such polypeptides making up the protein may be linked by disulfide bonds. The term "protein" may also comprehend a family of polypeptides having identical amino acid sequences but different post-translational modifications,

10   particularly as may be added when such proteins are expressed in eukaryotic hosts.

As used herein, the term "protein fractionation" refers to an analytical technique used to separate molecules. Several of the methods of fractionation well-known to those of skill in the art include chromatography, electrophoresis, and isoelectric focusing.

The invention provides a method for preparing a liquid sample for fractionation. As used

15   herein, the terms "sample" and "biological sample" are used interchangeably, and include material derived from an animal, preferably a mammal, more preferably a human. In preferred embodiments, a sample is a liquid sample derived from a human or animal. Such liquid samples include but are not limited to blood, plasma, serum, serum derivatives, bile, phlegm, saliva, sweat, tears, amniotic fluid, urine, peritoneal fluid, lymph, vaginal secretion, semen, spinal fluid,

20   ascitic fluid, saliva, sputum, breast exudate, synovial fluid and cerebrospinal fluid (CSF), such as lumbar or ventricular CSF.

The terms "detectable trait", "trait" and "phenotype" are used interchangeably herein and refer to any visible, detectable or otherwise measurable property of an organism. Typically the terms "detectable trait", "trait", or "phenotype" are used herein to refer to symptoms of, or

25   susceptibility to, a disease; or to refer to an individual's response to an agent, drug, or treatment acting on a disease; or to refer to symptoms of, or susceptibility to, side effects to an agent acting on a disease.

As used herein, "fraction" refers to part of a sample which has been split in a number of sub-samples of significantly differing compositions, typically as obtained after a chromatography

30   step.

As used herein, "aliquot" and "portion" refers to part of a sample which has been split in a number of sub-samples of equivalent compositions, in particular containing similar proteins in similar quantities. Samples can be aliquoted at any stage in the methods of the invention, e.g. to prevent overloading a column when processing the sample by conducting multiple runs on

aliquots is preferred. Similar fractions resulting from multiple runs may be pooled together at any stage in the methods of the invention. It will therefore be apparent to one skilled in the art that aliquot, or portion, as used herein, may represent any volume of sample, and is not limited to small sample volumes.

5

Biological samples

It is contemplated that the method can be used to identify markers, targets, or protein therapeutics in tissue or body fluid samples. The method, however, is particularly useful for the analysis of a body fluid, for example, blood, plasma, serum, serum derivatives, bile, phlegm,

10   saliva, sweat, tears, amniotic fluid, urine, peritoneal fluid, lymph, vaginal secretion, semen, spinal fluid, ascitic fluid, saliva, sputum, breast exudate, synovial fluid and cerebrospinal fluid (CSF), such as lumbar or ventricular CSF. Plasma, however, is most preferred.

The plasma samples are preferably maintained such that protein quality is not compromised. Conventional means of storing plasma samples can be used, such as addition of

15   protease inhibitors, and anti-coagulants.

In order to maintain the quality of the protein complement in the biological sample, the sample is obtained and stored using known methods for maintaining a sample. When the biological sample is blood or plasma, anticoagulant and antiprotease compositions are added to the blood or plasma samples. Especially when the sample is blood, blood cell activation,

20   aggregation and adhesion inhibitors are added. Commonly used anticoagulants fall into two general classes, the thrombin inhibitors and the calcium chelators. Of the thrombin inhibitors, heparin is the most commonly used today. It is a known inhibitor of acid phosphatase, lactate dehydrogenase, beta-hydroxybutyrate dehydrogenase, glutamyl transferase, creatine kinase, and restriction endonucleases. Of the calcium chelators, ethylenediaminetetraacetic acid (EDTA),

25   sodium citrate and oxalate salts are commonly used. Further anticoagulants are provided in WO 95/14788, the disclosure of which is incorporated herein. Commonly used and commercially available antiproteases include serine protease inhibitors (e.g., reversible and irreversible thrombin inactivators, Xa factors), inhibitors of cysteine proteases, calpain proteases and metalloproteases. Blood cell activation, aggregation and adhesion inhibitors include for example

30   platelet activation inhibitors, platelet-platelet interaction inhibitors, phospholipid-binding inhibitors, and fibrinogen inhibitors. In general, the anticoagulation and antiprotease mechanisms are selected so as not to interfere with methods for detecting proteins in the biological sample.

In general, a blood sample will be introduced into a means for receiving blood, such as a syringe or capsule, which has blood contacting surfaces that are coated with anticoagulant compositions. Anticoagulants as well as antiprotease compounds can also be added to the blood or plasma sample itself. Preferably, a combination of anticoagulants and/or antiprotease

5   compounds are used. For example, mixtures generally comprise at least two serine protease inhibitors in addition to the blood coagulation-retardant compounds, and in the case of blood also blood cell activation, aggregation and adhesion inhibitors.

It will be appreciated that any known procedures can be used to process and maintain biological samples. Several examples of samples preparation methods for use in proteomic

10   analyses are provided in Sanchez, J.-C. Practical aspects of 2-DE: Sample preparation and solubilization ABRF '98, San Diego (1998), the disclosure of which is incorporated herein by reference.

The methods of the invention are particularly advantageous since they can be used in connection with a wide range of scientific protocols, particularly using widely varying amounts of

15   fluid sample. In one example, separation is carried out at an industrial scale on a large quantity of sample, thereby providing highest sensitivity for the detection of rare proteins and peptides. In such examples, preferably at least 100 mL, 500mL, 1L, 2L, 10L, 20L, 25L or more of fluid sample are processed using the methods of the invention. Preferably the protein component, optionally excluding specifically removed proteins depending on the particular method used,

20   from at least 100 mL, 500ml, 1L, 2L, 10L, 20L, 25L of biological sample (for example blood, plasma, serum, urine, etc.) is introduced to a size exclusion chromatography column. As further exemplified herein, a large volume of sample (2.5L in the case of Example 1) can be divided into a plurality of aliquots and introduced to the column separately so that a plurality of runs are carried out. If desired, the components eluted from the size exclusion column can be pooled at

25   any suitable time thereafter during the separation process.

In other examples, the invention can be used in the separation and analysis of proteins and peptides using smaller volumes. This is more applicable in situations where less fluid sample is available such as in the case of CSF for example, or less sensitivity is required. In these aspects, preferably at least 1 mL, 5mL, 10mL or 50mL of fluid sample are processed using the

30   methods of the invention.

The samples may be obtained from any suitable number of individuals. In one aspect, a sample is obtained from a single individual. For example, a proteomics study may involve detecting proteins present in a sample from an individual displaying a given trait and an individual not displaying said trait. Preferably the individuals will be matched for factors other

than said trait. Any number of individuals can be used in a study. For example, the methods according to the invention (e.g. separation or fractionation of proteins, and detection) can be repeated for separate samples from at least 2, 5, 10, 20, 50, 75, 100 or 1000 individuals.

In other aspects, particularly where large volumes of fluid sample are used, the sample
5   will generally be a pooled sample that combines samples from a number of individuals. This method also allows comparison between samples from trait-displaying individuals and reference individuals, thereby resulting in an averaging of the levels for the proteins present in each pooled sample. Preferably a pooled biological sample contains biological fluid from at least 2, 5, 10, 20, 50, 75 or 100 individuals. Thus, an exemplary pooled plasma sample may comprise at least 10
10   mL, or preferably at least 50 mL of plasma from each of at least 50 individuals.

Particularly when using larger volumes of sample, it will be appreciated that a quantity of pooled sample to be processed according to the methods of the invention may be provided as a single sample, or may be divided into a plurality of samples. For example, a sample may be divided into a plurality of substantially identical smaller samples, herein described as "aliquots"
15   or "portions". This permits large volumes of biological samples to be treated without overloading a separation column. In one example, a series of sample runs can be combined after a separation step by combining analogous fractions. In one embodiment, analogous fractions are combined following the separation step, that is, without subjecting the fractions to a further separation step. In another example, the fractions obtained following a separation step are subjected to one of
20   further chromatographic separation steps, and only then are analogous fractions combined. The fractions that are combined can ultimately be analyzed using mass spectrometry techniques, increasing the ability of the analysis system of the invention to detect proteins present in plasma in low abundance. Further details and examples for aliquoting, and combining or pooling following separation steps are further provided below.

25

Liquid Chromatographic Separation

Chromatographic techniques are well known in the art as means for separating components (solutes) present in a mixture. These techniques are particularly useful in the chemical and biotechnological arts. True chromatography describes the separation of solutes
30   according to their different partitioning between two (or three) phases. The phases generally are solid and liquid, and solute partitioning results in their differing mobilities through a layer of solid, typically particulate, matrix in the presence of a flowing phase. Solute transfer through the layer may be along a pressure gradient, generally referred to as "liquid chromatography". Typically, the sample to be separated is applied to a column filled with pellets or grains of a

chromatography separation medium, and a solvent flow is maintained through the column at a steady rate. Components of the mixture are carried along by the solvent flow until each substance exits the column as a "peak" in the output, different peaks being more or less broad and overlapping.

5   Chromatographic matrices can separate components by any of a number of criteria, including size, electrical charge, hydrophobic interaction, and/or specific affinity for the matrix or binding sites thereon. Because the components in the mixture will vary in their affinity for the matrix, their partitioning as they pass through the matrix separates the components so that they exit the matrix sequentially, separated temporally and spatially. Determination of the location of

10   the various separated components, or of a given component of interest within the sequence, generally is achieved by collecting the fluid phase exiting the matrix (i.e., the effluent stream) as a series of fractions and sampling these fractions to identify their contents by any of a number of means known in the art.

15   *Removal of abundant proteins*

         In view of the large dynamic range of abundance of proteins in some mammalian samples, removal of abundant proteins provides improved sensitivity for the detection of low-abundance proteins. In preferred embodiments, means are used which remove known abundant proteins, thereby minimizing the undesired removal of proteins to be identified according to the

20   methods of the invention. One or a plurality of abundant proteins or class of abundant proteins may be removed. Preferred methods use affinity chromatography, although it will be appreciated that the invention is not limited thereto, and any suitable non-chromatographic means can be used to remove abundant proteins.

         The presence of abundant proteins can present significant limitations on the detection of

25   proteins in a sample. For example, a single protein, albumin, makes up more than 50% of the total human serum protein content by weight (Putnam, R.W., *The Plasma Proteins*, Academic Press, New York 1975). Also present at high concentrations in plasma are immunoglobulin (Ig) light and heavy chains. The presence of the abundant albumin and Ig impedes effective resolution and detection of many low abundance proteins. Thus, in preferred aspects, to improve liquid

30   chromatographic fractionation and 2-D electrophoresis of plasma, human serum albumin and/or immunoglobulins are specifically removed. In other preferred aspects, abundant proteins making up at least 5%, 10%, 25%, 50%, 60%, or 75% (w/w) of the total protein mass of a liquid sample are removed using means for the specific removal of abundant proteins. Among the main proteins that can be removed from plasma are haptoglobin and transferin.

Immunoglobulins and/or albumin proteins can be extracted using any suitable conventional methodologies used in the art. In a preferred aspect, affinity-based methodologies are used. For example, immunoglobulin can be removed selectively from samples using binding proteins, for example, an antibody or a fragment thereof, Protein A, or Protein G, immobilized on

5   a solid support. For example, a solution of interest can be passed through a chromatography column packed with such a solid support under conditions such that the immunoglobulin molecules preferentially bind to the matrix. The resulting column flow through, therefore, is depleted of immunoglobulin. Similar methods can be used with albumin ligand to remove serum albumin.

10   In preferred embodiments, a fluid sample comprising substantially undigested proteins and peptides, e.g. which has not been substantially subjected to any digestion process, is provided. Additionally, when the step for removal of abundant proteins is carried out prior to the size-exclusion chromatography step, the fluid sample has preferably not been subjected to any filtration procedure that would remove a substantial number of proteins or peptides having a

15   molecular weight of less than about 20kDa from the sample. Referring to FIG 1, the sample is injected onto two inline affinity chromatography columns, one column comprising an albumin ligand and a second column comprising Protein G, shown in step 101. The non-retained fraction can then be processed further immediately according to the invention, such as by collecting fractions and injecting to another column (e.g. size exclusion) or by processing on-line on a

20   further column. Alternatively, the non-retained fraction, or a portion thereof, can be frozen and stored for further processing at a later time. When using large quantities of sample, the sample is divided into several aliquots and multiple runs can be performed, such as 20 runs as shown in FIG 1. Non-retained fractions from multiple runs can either be pooled, or can be processed further individually. In a preferred example in FIG 1, fractions are frozen and stored individually and

25   then injected onto a size exclusion column.

It will be appreciated that various suitable matrices can be used. For example, Protein G coupled to agarose particles, available commercially from Amersham, Upsala, Sweden. Similarly, albumin can be removed selectively for samples of interest via affinity chromatography, using, for example, Sepharose coupled to Cibacron blue available commercially

30   from Pharmacia and Upjohn, Peapack, NJ, or using an albumin ligand, available commercially from Amersham, Upsala, Sweden. Alternatively, both albumin and immunoglobulin G can be removed simultaneously from serum using techniques such as those described in Lollo *et al* ((1999) Electrophoresis 20:854-859).

17

In most preferred methods, abundant proteins are removed using highly specific ligands. Examples of preferred ligands are provided in International Patent Publication No. WO 00/37501, the disclosure of which is incorporated herein by reference. Although removal of serum albumin using Cibacron-blue can be used, more preferred methods use specific ligands that remove

5   substantially only the protein of interest. Cibacron-blue dye binds many proteins other than albumin, such as interferon, lipoproteins, blood coagulation factors, kinases, dehydrogenases and most enzymes requiring adenyl-containing cofactors. On the other hand, a specific ligand typically exhibits an affinity between the reagent and the specific protein to be removed of at least about 10^6 M Ka.

10   Specific ligands may be generated *de novo* against proteins to be removed or selected from libraries of ligands. Examples of specific ligands include antibodies (e.g., polyclonals from animals, monoclonals from hybridomas), single chain antibodies from phage display libraries (Vaughan *et al.*, *Nat Biotechnol.* 1996, 14:309-14), small peptides (Norman *et al. Science* 1999, 285:591-5) or RNA-protein fusion libraries (Kreider *et al. Med Res Rev* 2000, 20:212-5), DNA

15   and RNA aptamers (Kusser *et al. J Biotechnol.*, 2000. Mar; 74:27), small molecules (Macbeath *et al. J. Am. Chem.* Soc. 1999, 121:7967-7968), random length peptides and proteins (Walter *et al. Curr Opin Microbiol.*, 2000. 3:298-302), as well as natural or recombinant receptor proteins, and their fragments (Alexander and Peters, 2000, *Trends in Pharmacological Sciences*, Published by Current Trends, London).

20   In a preferred embodiment, a liquid sample is depleted from abundant proteins by contacting the sample with a polypeptide affinity reagent. As used herein, the terms "polypeptide affinity reagent" refers to a polypeptide that specifically binds to a macromolecule of interest in a liquid sample to be depleted. "Specifically binds" means the adsorptive polypeptide recognizes and binds a specified macromolecule, but does not substantially recognize and bind other

25   molecules in a sample, e.g., a liquid biological sample, that naturally includes a variety of macromolecules. The principle is to contact the liquid sample with reagents having specific affinity for a particular component or defined class of components. These reagents have narrow specificities for particular sets of macromolecules.

Antibodies represent the main class of polypeptide affinity reagents that are

30   immunoreactive or bind to epitopes of macromolecules. The term "epitope" refers to any antigenic determinant on an antigen to which an antibody binds. Epitopes usually are chemically active surface groupings of amino acids or sugar side chains and usually have specific three-dimensional structural characteristics, as well as specific charge characteristics.

As used herein, the term "antibody" includes intact antibody molecules as well as fragments thereof, such as Fab, Fab', Fv, and single chain antibody that can bind the epitope. These antibody fragments retain some selective ability to bind the corresponding antigen or receptor. Particularly useful antibodies include polyclonal and monoclonal antibodies, chimeric

5   antibodies, single chain antibodies and the like, having the ability to bind with high immunospecificity to abundant macromolecules.

The preparation of polyclonal antibodies is well-known to those skilled in the art. See, for example, Green et al. ("Production of Polyclonal Antisera", in Immunochemical Protocols, Manson, ed., Humana Press, 1992, pages 1-5) and Colligan et al. (Production of Polyclonal

10 Antisera in Rabbits, Rats, Mice and Hamsters, in: Current Protocols in Immunology, section 2 1, 1992), incorporated herein by reference. The preparation of monoclonal antibodies likewise is conventional. Monoclonal antibodies can be produced using methods well known in the art. See, Kohler et al. (Nature 256: 495, 1975); Current Protocols in Molecular Biology (Ausubel et al., ed., 1989); and Harlow and Lane (Antibodies: A Laboratory Manual, Cold Spring Harbor

15 Laboratory, New York, current edition), incorporated herein by reference. Briefly, monoclonal antibodies can be obtained by injecting mice with an antigenic composition, verifying the presence of antibody production by removing a serum sample, removing the spleen to obtain B lymphocytes, fusing the B lymphocytes with myeloma cells to produce hybridomas, cloning the hybridomas, selecting positive clones that produce antibodies to the antigen, and isolating the

20 antibodies from the hybridoma cultures. Monoclonal antibodies can be isolated and purified from hybridoma cultures by a variety of well-established techniques. Such isolation techniques include affinity chromatography with protein-A Sepharose, size-exclusion chromatography, and ion-exchange chromatography.

Methods of *in vitro* and *in vivo* multiplication of monoclonal antibodies are well known

25 to those skilled in the art. Multiplication *in vitro* may be carried out in suitable culture-media such as Dulbecco's Modified Eagle Medium (DMEM) or R.PMI 1640 medium, optionally replenished by a mammalian serum such as fetal calf serum or trace elements and growth-sustaining supplements such as normal mouse peritoneal exudate cells, spleen cells, bone marrow macrophages. Production *in vitro* provides relatively pure antibody preparations and allows scale-

30 up to yield large amounts of the desired antibodies. Large scale hybridoma cultivation can be carried out by homogenous suspension culture in an airlift reactor, in a continuous stirrer reactor, or in immobilized or entrapped cell culture.

Multiplication *in vivo* may be carried out by injecting cell clones into mammals histocompatible with the parent cells, e.g., syngeneic mice, to cause growth of antibody

producing tumors. Optionally, the animals are primed with a hydrocarbon, especially oils such as pristane (2,6,10,14-tetramethyl-pentadecane) prior to injection. After one to three weeks, the desired monoclonal antibody is recovered from the body fluid of the animal.

5 If desired, polyclonal or monoclonal antibodies can be further purified, for example, by binding to and elution from a matrix to which the polypeptide or a peptide to which the antibodies were raised is bound. A purified antibody may be obtained, for example, by affinity chromatography using recombinantly-produced protein or conserved motif peptides and standard techniques. Those of skill in the art will know of various techniques common in the immunology arts for purification or concentration of polyclonal antibodies, as well as monoclonal antibodies.

10 See, e.g., Colligan, et al. (Unit 9, Current Protocols in Immunology, Wiley Interscience, 1997).

As used herein, the term "albumin-specific monoclonal antibodies" refers to monoclonal antibodies that specifically bind to serum albumin. "Specifically binds to albumin" means the monoclonal antibody recognizes and binds to serum albumin, but does not substantially recognize and bind other molecules in a sample, e.g., serum or plasma, that naturally includes serum

15 albumin. The invention provides a monoclonal antibody that can immunoprecipitate serum albumin from serum or plasma. This means that the monoclonal antibody recognizes an epitope on the albumin molecule that is not blocked by the numerous plasma and serum proteins that bind to albumin in serum or plasma.

Although polyclonal antibodies provide specificity, there is the inherent variability in

20 antibody population that occurs during separate immunization schedules that can lead to reproducibility problems. Additionally, the supply of polyclonal antibody containing serum is limited by the health and finite lifespan of the producing animal. Considering the large quantities of antibody required for the treatment of plasma, the use of polyclonal antibodies is possible but not preferred.

25 Preferred ligands for the removal of abundant proteins include non-antibody peptide and protein ligands. Various commercially available peptide or protein ligands can be used. It will be appreciated that peptide ligands may also be prepared using known methods for generating and testing combinatorial peptide libraries.

Examples of polypeptide affinity reagents include protein A and protein G. Protein A is a

30 protein of MW 42,000 from the bacterium Staphylococcus aureus that binds to IgG from a wide range of species, including human, rabbit, donkey, pig, and guinea-pig. Protein A is commonly used as a secondary reagent in immunological and biological techniques, as described by Goding (J. Immunol Meth. 20: 241-253, 1978), and is available from Scripps Laboratories. Protein G is a monomeric protein (MW 63,000) from human group G streptococcus. Protein G possesses two or

three antibody-binding sites and binds IgG from a wide range of species. Compared to protein A, protein G binds with a higher affinity to rat, mouse and goat IgG, as described by Bjrk et al. (J. Immunol 133, 969-974, 1984).

5   In another example, a macromolecule-polypeptide affinity reagent complex is formed, that is then contacted with the other member of a high affinity binding pair system to form a "collapsible affinity matrix" (as described in PCT patent publication No. WO 99/39204). The collapsible affinity matrix is specific for the abundant macromolecule and, when centrifuged, contains very little dead volume that would-otherwise trap additional sample macromolecules. For example, biotinylated adsorptive proteins can be used in such a method. In a specific

10   embodiment, a biotinylated anti-Human Serum Albumin (HSA) antibody, in conjunction with avidin and human serum, forms a collapsible affinity matrix, containing albumin. The combination of biotinylated protein A, avidin, and human plasma, followed by contact with biotinylated anti-HSA and avidin allows simultaneous co-precipitation of albumin and immunoglobulin (Ig). The practice of the method of the invention can thereby provide plasma

15   samples substantially depleted of albumin and immunoglobulin.

Methods for removal of specific abundant proteins can be carried out one or a plurality of times on a single liquid sample to remove multiple abundant molecules.

The invention thus preferably includes preparing a liquid sample according to the invention for fractionation by contacting the liquid sample with an affinity reagent having

20   specificity for an abundant macromolecule in the sample. Examples of macromolecules include polynucleotides, polypeptides, and polysaccharides. Examples also include glycoproteins, in which saccharide (sugar) moieties are covalently bound to polypeptides, and nucleoproteins, which are complexes of polynucleotide and polypeptide. An "abundant macromolecule" or an "abundant protein" is respectively a macromolecule or protein present in a sample in such

25   quantity that the presence of the macromolecule or protein interferes with an aspect of the analysis of the sample. An abundant macromolecule or protein may constitute for example at least about 1%, 5%, 10%, 25% or 50% of the total protein in a sample. In most preferred aspects, a plasma or serum sample is obtained that is substantially depleted in albumin and immunoglobulins. As used herein, the term "substantially depleted" means that the most abundant

30   proteins have been removed such that the plasma or serum sample contains less than 50% (w/w) of the total proteins before depletion.

Size exclusion of intact proteins and peptides

The fluid sample is subjected to at least one size-exclusion separation or fractionation step. Although other methods such as ultrafiltration methods can be used, use of size-exclusion liquid chromatography is preferable due to advantages, e.g. in protein recovery yield.

5   Again, since the present invention provides a system for analyzing the biomolecules present in a biological sample, typically the proteins and peptides present in a "proteome", it will be appreciated that the samples to be injected to the size exclusion column are not pre-treated in such a way as to remove a substantial number of proteins to be identified. In preferred embodiments, a sample will generally not be subjected to a treatment which would substantially

10   reduce the diversity of the protein or peptide complement of the sample to be analyzed according to the methods of the invention, or which would non-specifically remove proteins or peptides from the sample based on criteria other than size. For example, a plasma sample will preferably not be subjected to a conventional ultrafiltration step which is known to non-specifically remove proteins which may stick to the ultrafiltration membrane. However, it will be appreciated that the

15   sample may be aliquoted so long as substantially all protein-containing aliquots obtained thereby are subsequently included in the size-exclusion separation step. Likewise, the biological sample will not have been treated so as to cause the digestion of a substantial number or amount of proteins and peptides in the sample. Preferably, the sample has not been treated by the addition of trypsin or other proteolytic enzyme prior to the size exclusion step.

20   As shown in Figure 1, non-retained fractions stored following affinity chromatography for the removal of abundant proteins are injected onto two inline gel filtration columns (103). Each fraction from affinity chromatography step (101) is injected separately to the gel filtration columns (103) resulting in a total of 20 runs. The second of the gel filtration columns is connected inline to a reverse phase capture column, as further discussed below. Eluted fractions

25   enriched in small (e.g., 20kD or lower) proteins and peptides can be retained for further processing.

Thus, as used herein, the fractions which are not enriched in small proteins, that is, the fractions containing the bulk of the large proteins (above 20 kD), need not be retained for further processing and in particular need not be introduced into the next separation steps of the invention.

30   Thus, advantageously, the size exclusion step can be a preparative step, which means that it is used solely to remove large proteins from the sample and not to resolve the complexity of the proteins contained in the sample by, e.g. fractionating the sample into a large number of fractions which would be all individually further processed by the methods of the invention.

Moreover, in preferred embodiments of the methods, all the output fractions from said size exclusion column which comprise, or are enriched in, proteins of less than 20kD are introduced to a further chromatography column. Treating the complete set of proteins under the molecular weight cutoff provides a more complete inventory of low molecular weight proteins from a proteome. Thus, as used herein, "retaining the proteins of less than 20kD" means retaining substantially all proteins from the sample which have a molecular weight below 20kD, with the understanding that this may be somewhat limited by technical constraints. In these embodiments of the invention, more than 80%, preferably more than 85 %, even more preferably more than 90 %, even more preferably more than 95 %, even more preferably more than 98 %, still more preferably more than 99 % of the proteins of less than 20kD contained in the initial sample are retained by the size exclusion step and processed further by the methods of the invention.

Size exclusion chromatography, also known as gel permeation or gel filtration chromatography, does not involve any adsorption and has the advantage of being fast. The packing is a porous gel, and is capable of separating large molecules from smaller ones. The larger molecules elute first since they cannot penetrate the pores. This method is common in protein separation and purification processes.

Preferably, at least 1 mL, 5 mL, 10 mL, 50 mL, 100 mL, 200mL, 500 mL, 750 mL, 1 L, 2 L, 5 L or 10 L of fluid sample comprising substantially intact and undigested proteins will be subjected to the size-exclusion based separation or fractionation step. It will be appreciated that where said sample has already been separated into multiple aliquots, the aliquots can be passed over the gel filtration column during the course of a plurality of runs.

It will be appreciated that any suitable filtration device, medium and conditions can be used in accordance with this aspect of the invention. When large volumes of sample are used, the sample can be injected to the gel filtration column(s) in a series of aliquots. Thus, at least 2, 4, 8, 10 or more gel filtration runs can be carried out. Where large volumes of sample are used, the gel filtration column has a volume of at least 10 mL, 100 mL, 200 mL, 300mL, 400 mL, 500 mL, 750 mL, 1 L, 2 L, 5 L or 10 L of fluid sample. Alternatively, two in line columns may be used.

In one aspect of the invention, column volume is comprised between 10 and 40 liters, with column diameter comprised between 10 and 30 cm. The column size is defined according to the sample volume so that the sample volume does not exceed 5% of column volume. Linear velocity is comprised between 15 and 60 cm/h, depending on the media and apparatus used. A large range of media can be used; the fractionation range must ensure an adapted separation for the chosen molecular weight cutoff, e.g. 20 kDa. For an extensive description of available media,

see "Size exclusion chromatography", M. Rögner, in "Protein liquid chromatography", M. Kastner Ed., 2000, Elsevier. Protein elution is monitored by 280 nm UV absorption and cut off for <20 kD protein collection is realized at a defined chromatogram absorbance value, depending of the nature of biological fluid. SDS-PAGE may also be used to monitor the effectiveness of the

5   size exclusion column. Composition of chromatography buffer is critical since it is important to prevent low molecular weight proteins and peptides adsorption to major plasma protein. The choice of pH and ionic strength, and their combination, is critical to preserve an optimal separation. Urea is optimally added to this buffer to prevent specific or nonspecific protein molecules aggregation. As an example, approximately 85% of the retained proteins have a

10   molecular weight below 20kD in the protocol described in Example 1.

Reverse Phase Capture

Following gel filtration, the protein and peptide samples are preferably concentrated. A preferred method for concentrating proteins and peptides is to use a reverse phase capture

15   column. The reverse phase capture column(s) (103) can be advantageously connected inline to the gel filtration column. In embodiments where large volumes are used and/or where lower capacity columns are used, multiple gel filtration aliquots are processed on the inline reverse phase capture columns, leading in turn to multiple reverse phase capture aliquots. These fractions can then be stored or further processed individually or may be pooled according to any desired

20   arrangement. In an exemplary protocol shown in Figure 1, fractions obtained from a preceding inline gel filtration column are further processed on two inline reverse phase capture columns. Eluted fractions are pooled and further split into 6 aliquots of about 200 mg of protein (105).

It will nevertheless be appreciated that any other suitable concentration method can be used following gel filtration. Alternatively, if adequate protein concentration can be achieved

25   using size exclusion chromatography, no concentration step is needed.

Separation according to many dimensions

The analysis of complex mixtures according to the invention preferably involves separation according to several dimensions in order to resolve all the components present in a

30   sample. It is for this reason that multi-dimensional separation schemes have been devised. When constructing a successful multi-dimensional system several criteria need to be addressed. In one aspect, the techniques should base their respective separations on widely variant criteria. Doing so will reduce the amount of redundant information contained in the resulting dataset. Thus, in addition to, and most preferably subsequently to, size-exclusion based separation, a biological

sample is treated according to one or more separation protocols as described herein. Most preferably, substantially the entire protein content, and thus the protein diversity, or substantially all the retained protein-containing fractions from the size-exclusion step are subjected to further separation steps.

5   In preferred aspects, the invention discloses a method for identification of proteins comprising a multidimensional liquid chromatography separation process. Following a size-exclusion-based step and optional concentration step(s), proteins and peptides are subjected to one or a plurality of liquid chromatographic separation or fractionation steps. This process is preferably carried out on protein fractions obtained from said size-exclusion step which are

10   enriched in low molecular weight proteins, preferably proteins of less than 20kD.

Thus, in one aspect, fractions enriched in proteins of lower molecular weight, preferably enriched in proteins of less than 20kD, are subjected to liquid chromatography protocols as further described herein. Preferably liquid chromatography protocols are selected from the group consisting of ion exchange, reverse phase, affinity chromatography, Immobilized Metal Affinity

15   Chromatography, Dye-Ligand affinity chromatography, Hydrophobic Interaction Chromatography, hydroxylapathite chromatography, chromatofocusing. Details and preferred embodiments are further described herein.

Ion (cation) exchange chromatography

20   In preferred embodiments, fractions obtained via size exclusion chromatography are subjected to ion (e.g. cation or anion) exchange chromatography, preferably cation exchange, followed by reverse phase HPLC.

In preferred embodiments, fractions enriched in proteins and peptides of 20kD or less are obtained from gel filtration and a subsequent concentration step, and are injected to a cation

25   exchange column (106).

Particularly where large volumes of samples are used, multiple runs (e.g. from individual aliquots following size exclusion) on cation exchange are carried out, such that a plurality of fractions are obtained for each run. In one example, 6 runs are carried out for the 6 aliquots obtained from reverse phase capture, each containing about 200 mg of protein, and each run

30   resulting in about 20 fractions from cation exchange. Similar fractions from different runs are then pooled together.

Ion exchange chromatography is commonly used in the purification of biological materials. There are two types of ion exchange: cation exchange in which the stationary phase carries a negative charge, and anion exchange in which the stationary phase carries a positive

charge. Charged molecules in the liquid phase pass through the column until a binding site in the stationary phase appears. The molecule will not elute from the column until a solution of varying pH or ionic strength is passed through it. Separation by this method is highly selective. Since the resins are fairly inexpensive and high capacities can be used, this method of separation is applied

5   early in the overall process. The choice of pH must ensure quantitative adsorption of the proteins. The ionic strength of the buffer used must be low for the same reason. Chromatography with cation exchange media needs the use of acidic pH, compatible with biological fluid to prevent protein precipitation. Chromatography with anion exchange media needs the use of basic pH, at a value compatible with preservation of the peptide link. The column size is defined according the

10   protein quantity to fractionate, determined after capacity studies. The column height preferably does not exceed 25 cm and the column diameter is deduced from capacity and column height according the following equation:

Column Volume = (π x Column Radius^2) x Column Height,

with Column Volume = (Protein Amount / Capacity) x 2

15   Linear velocity with actual media may reach 300 cm/h but optimal value must be carefully determined to prevent loss of resolution. Protein elution is monitored by 280 nm UV absorption. The mobile phase is composed of two solutions. Using the example of cation exchange, the first solution (A) is composed of an acidic buffer as further provided. The second solution (B) is composed of the same buffer as in A containing a high concentration of salt (e.g.,

20   NaCl 1M). The column is equilibrated with 100 % of solution A, the protein sample is injected and the column is washed with solution A. Adsorbed proteins are progressively eluted and so resolved by a linear or step gradient from 0 % of B to 100% of B. The elution profile of a given polypeptide from cation exchange is linked to the number of positive charges.

For a review, see "Ion Exchange Chromatography", H. Roos, in "Protein liquid

25   chromatography", M. Kastner Ed., 2000, Elsevier, the disclosure of which is incorporated herein by reference.

Eluted protein solution must be compatible with the following step, in one example reverse phase high performance liquid chromatography (RP-HPLC).

Some of the fractions obtained from cation exchange may or may not be pooled,

30   depending on the needs of the user. In an exemplary case described in Figure 1, 6 runs are carried out on a cation exchange column, resulting in 20 fractions for each run. These 120 fractions are then reduced to 20 fractions by combining fractions from each of the 6 runs. In Example 1, another pooling strategy is described. It will be appreciated that pooling of eluted

fractions may be inter-run or intra-run, and fractions to be combined may be selected according to any suitable criteria. Preferably analogous samples from multiple runs are pooled.

Reverse Phase HPLC

In preferred embodiments, fluid samples are fractionated on a reverse phase column (111 and/or 116). Reverse phase chromatography is a powerful analytical tool and involves a hydrophobic, low polarity stationary phase which is chemically bonded to an inert solid such as silica. The separation is essentially an extraction operation and is useful for separating non-volatile components. In preferred multidimensional fractionation schemes of the invention, both reverse phase chromatography and ion exchange chromatography may be used, in any order.

Most preferably, the method of the invention combines ion exchange and reverse phase HPLC. In a preferred example, samples are injected onto an ion exchange, preferably cation exchange, column (106), and subsequently fractions eluted therefrom are injected onto a reverse phase column (111). As shown in Figure 1, following ion exchange chromatography (106), eluted fractions may be reduced and alkylated (108) before injection to a reverse phase column (111). If desired, a reverse phase column (111) may be connected inline to the ion exchange column (106).

In other configurations of said preferred arrangement where ion exchange and reverse phase chromatography are used, and moreover particularly when using larger volumes of fluid sample, eluted fractions from the ion exchange column are collected and fractions are pooled (107) prior to injection of the fractions to the reverse phase High Pressure Liquid Chromatography (HPLC) column (111). As discussed, pooling (107) may be inter-run or intra-run, and fractions to be combined may be selected according to any suitable criteria.

Reverse phase HPLC can be configured to produce a number of fractions as desired. In the exemplary protocol of Figure 1, 30 fractions are generated for each of the 20 fractions injected to the reverse phase column. Following reverse phase HPLC, fractions can be dried (115) for storage if desired.

In preferred embodiments, two or more sequential fractionations using reverse phase chromatography are carried out (111 and 116). Thus, in preferred aspects, a system of the invention comprises a first and a second reverse phase HPLC columns. Optionally, the second reverse phase column can be connected inline to said first column. In preferred embodiments, the protein concentration of one or a plurality of fractions is adjusted; for example fractions having

27

an abnormally high protein level are diluted or injected as multiple runs, based on optical density (OD) measurements (109, 110, 113 and 114). Aliquots injected as multiple runs may be pooled back together at any later stage (112). Following the first reverse phase step, fractions can be dried (115) for storage if desired, and resuspended when a further fractionation (e.g. second

5   reverse phase fractionation) is to be carried out. Fractions are then injected to a second reverse phase HPLC column (116). In the exemplary protocol of Figure 1, for each of the 600 fractions generated following the first reverse phase column, the second reverse phase HPLC step generates 24 further fractions.

       Reverse phase high performance liquid chromatography (RP-HPLC) is similar to reverse

10   phase, only in this method, the process is conducted at a high velocity and pressure drop. The column is shorter and has a small diameter, but it is equivalent to possessing a large number of equilibrium stages. This method is essentially characterized by the use of low diameter chromatography beads, typically from 3 μm to 20 μm. The column dimensions are chosen as for ion exchange chromatography. The mobile phase is composed of two solutions. The first (A) is

15   composed of water acidified with trifluoroacetic acid for example (0.1% w/v). The second solution (B) is composed of solvent such as acetonitrile, containing a minor proportion of water (e.g. 20 %). The column is equilibrated with 100 % of solution A, the protein sample from ion exchange is injected and the column is washed with solution A. Adsorbed proteins are progressively eluted and so resolved by a linear gradient from 0 % of B to 100 % of B. Protein

20   elution is examined by UV adsorption at 210 nm. This technique requires the use of adapted pumps and detector since the pressure observed is high, commonly around 100 bars. For a review, see "Reversed-Phase Chromatography", H. Schlüter, in "Protein liquid chromatography", M. Kastner Ed., 2000, Elsevier.

Other protocols

25   In other aspects, it will be appreciated that other protocols to separate proteins can be used in addition to, or place of, the ion exchange- and/or reverse phase- based methods. A range of suitable multi-dimension protocols can thus be devised based in the present invention. Chromatography methods that may be used in such embodiments at some point in the separation process, preferably following the size exclusion step, include for example affinity

30   chromatography, immobilized metal affinity chromatography, dye-ligand affinity chromatography, hydrophobic interaction chromatography, hydroxylapathite chromatography, and chromatofocusing.

Affinity chromatography involves the use of a packing material which has been chemically modified by attaching a compound with a specific affinity for desired molecules, primarily biological compounds. The packing material used, called the affinity matrix, must be inert and easily modified. Agarose is the most common substance used. The ligands, or "affinity

5   tails", that are inserted into the matrix can be genetically engineered to possess a specific affinity. In a process similar to ion exchange chromatography, the desired molecules adsorb to the ligands on the matrix until a solution of high salt concentration or low pH is passed through the column (elution). Elution causes desorption of the molecules from the ligands, and they are released from the column. For a review on affinity chromatography technique, see "Protein purification,

10   principles and practice", R.K. Scopes Ed., 1994, Springer Verlag.

Immobilized metal affinity chromatography involves retaining proteins and peptides via accessible Trp and His residues. The matrix is attached to a metal chelator, and the chelated metal interacts with Trp and His via a non-covalent bond. The more common used metals are Cu++, Ni++, Zn++. Retained proteins are eluted by pH modification, EDTA metal complexing or,

15   preferably, using a competition ligand gradient. The more common competition ligand is imidazol. For a review, see "Immobilized Metal Ion Affinity Chromatography", M. Kastner, in "Protein liquid chromatography", M. Kastner Ed., 2000, p 301-383, Elsevier.

Dye-ligand affinity chromatography involves the use of immobilized dye linked to a chromatography matrix. The dye ligand, mainly polyaromatic sulfonated dyes, are able to replace

20   natural coenzyme ligands, ensuring adsorption of proteins containing such coenzymes. Some dye-ligand matrices are commercially available. For an extensive list of media and chromatography conditions, see "Dye-Ligand Affinity Chromatography" J. Kirchberger and H-J. Böhme, in "Protein liquid chromatography", M. Kastner Ed., 2000, p 415-448, Elsevier.

Hydrophobic interaction chromatography involves interaction between hydrophobic

25   surface areas of proteins and a mild hydrophobic ligand, immobilized on a chromatography media. Protein adsorption is favored by high concentrations of salts and elution is performed with a negative salt gradient, from higher to lower salts concentration. Preferred salts are ammonium sulfate, sodium sulfate or sodium chloride. It is also possible to use solvents for highly hydrophobic proteins elution. For an extensive list of media and chromatography conditions, see

30   "Hydrophobic Interaction chromatography", L.R. Jacob, in "Protein liquid chromatography", M. Kastner Ed., 2000, p 235-269, Elsevier.

Hydroxylapathite chromatography, in contrast to other chromatography technique, is based on interaction with a mineral matrix. The media is a crystalline form of calcium phosphate, which adsorbs proteins on their surface. The separation mechanism involved is not well

understood although the matrix has positive (Ca++) and negative charges (PO₄³⁻). For a review, see " Hydroxylapathite chromatography", W.R. Deppert and R. Lukacin, in "Protein liquid chromatography", M. Kastner Ed., 2000, p 271-299, Elsevier.

Chromatofocusing ensures the separation of proteins according to their pI. The media,
5   particular ion exchange media optimized for this purpose, binds amphoteric molecules (ampholines). The distribution of ampholines induces a pH gradient in the column and the separation principle is near those of electrofocusing, without the need of an electric field. This technique requires special media and buffer solutions. For a review, see " Chromatofocusing", R. Lukacin and W.R. Deppert, in "Protein liquid chromatography", M. Kastner Ed., 2000, p 385-
10   414, Elsevier.

To achieve a particular separation, the general practice of chromatographic separation involves identifying or selecting a particular medium or coated medium, and an optimum solvent, solvent flow rate, pH, ionic concentrations and other environmental conditions, such that the starting mixture will separate into a number of relatively narrow bands and such that
15   biomolecules (e.g. proteins) or mixtures of biomolecules sharing certain physical properties pass, or may be made to pass, as a distinct output.

The determination of an appropriate set of separation conditions for a particular biological fluid sample, which may have an as yet undetermined dynamic range and protein content, can be optimized or selected such that an appropriate chromatographic separation is
20   achieved. This system is able to separate biomolecules into large numbers of distinct fractions with minimal overlap of biomolecules between the fractions. The system is also adapted to allow the separation of low abundance biomolecules, including proteins present at subpicomolar concentrations in a fluid sample.

In general, the transport of material in a separation column proceeds on a macroscopic
25   level by flow past and between the grains or pellets of the chromatography medium, while the degree of separation and column capacity are governed more by the rates at which the particular components diffuse along branching paths into and out of pores in the medium, and are repeatedly adsorbed and released along the diffusion path. By increasing the flow rate to increase process output, one generally broadens the eluted peak width of each component, thus sacrificing
30   the resolution and hence the purity of the separated components; above a certain flow rate threshold, premature solute breakthrough may occur.

Status monitoring is particularly important in multistep preparation protocols. Monitoring of protein fractions can be used as part of a feedback system to adjust separation parameters.

Generally, a selected characteristic is determined using a previously established criterion for identification, for example, a characteristic absorbance measured at a given wavelength.

Within the methods of the invention, it will be apparent to one skilled in the art that any step may be run in parallel on a number of similar columns. This can be used for example to

5   process large samples with a high throughput, by removing the need, for aliquots of a sample, to run them one after the other on the same apparatus. They can be run simultaneously in parallel in a number of apparati similar to each other.

It is an object of the invention to provide a rapid and adaptable system and apparatus delivering reproducible results for identifying the presence and/or location of a molecule of

10   interest during any preparative or analytic protocol. The ultimate goal is to separate components of a protein mixture by exploiting the benefits of chromatographic techniques. Objects of the invention include multi-dimensional analysis to enhance resolving power of a chromatographic system coupled to means for the detection and identification of proteins in complex mixtures of body fluids.

15

Protein identification: MS and MS/MS

In accordance with the present invention, any instrument, method, process, etc. can be utilized to determine the identity of proteins in a sample. A preferred method of obtaining identity is by mass spectrometry, where protein molecules in a sample are ionized and then the resultant

20   mass and charge of the protein ions are detected and determined.

To use mass spectrometry to analyze proteins, it is preferred that the protein be converted to a gas-ion phase. Various methods of protein ionization are useful, including, e.g., fast ion bombardment (FAB), plasma desorption, laser desorption, thermal desorption, preferably, electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI). Many

25   different mass analyzers are available for peptide and protein analysis, including, but not limited to, Time-of-Flight (TOF), ion trap (ITMS), Fourier transform ion cyclotron (FTMS), quadrupole ion trap, and sector (electric and/or magnetic) spectrometers. See, e.g., U.S. Pat. No. 5,572,025 for an ion-trap MS.

Mass analyzers can be used alone, or in combination with other mass analyzers in tandem

30   mass spectrometers. In the latter case, a first mass analyzer can be use to separate the protein ions (precursor ion) from each other and determine the molecular weights of the various protein constituents in the sample. A second mass analyzer can be used to analyze each separated constituents, e.g., by fragmenting the precursor ions into product ions by using, e.g. an inert gas.

Any desired combination of mass analyzers can be used, including, e.g., triple quadrupoles, tandem time-of-flights, ion traps, and/or combinations thereof.

Different kinds of detectors can be used to detect the protein ions. For example, destructive detectors can be utilized, such as ion electron multipliers or cryogenic detectors (e.g.,
5   U.S. Pat. No. 5,640,010). Additionally, non-destructive detectors can be used, such as ion traps which are used as ion current pick-up devices in quadrupole ion trap mass analyzers or FTMS.

For MALDI-TOF, a number of sample preparation methods can be utilized including, dried droplet (Karasand Hillenkamp, Anal. Chem., 60:2299-2301, 1988), vacuum-drying (Winberger et al., In Proceedings of the 41st ASMS Conference on Mass Spectrometry and Allied
10   Topics, San Francisco, May 31-June 4, 1993, pp. 775a-b), crush crystals (Xiang et al., Rapid Comm. Mass Spectrom., 8:199-204,1994), slow crystal growing (Xiang et al., Org. Mass Spectrom, 28:1424-1429, 1993); active film (Mock et al., Rapid Comm. Mass Spectrom.,6:233-238, 1992; Bai et al., Anal. Chem., 66:3423-3430, 1994), pneumatic spray (Kochling et al., Proceedings of the 43rd ASMS Conference on Mass Spectrometry and Allied Topics; Atlanta,
15   GA, May 21-26, 1995, p1225); electrospray (Hensel et al., Proceedings of the 43rd ASMS Conference on Mass Spectrometry and Allied Topics; Atlanta, GA, May 21 -26, 1995, p947); fast solvent evaporation (Vorm et al., Anal. Chem., 66:3281-3287, 1994); sandwich (Li et al., J. Am. Chem. Soc., 11 8:11662-11663,1996); and two-layer methods (Dal et al., Anal. Chem., 71:1087-1091, 1999). See also, e.g., Liang et al., Rapid Commun. Mass Spectrom., 10: 1219-1226, 1996;
20   van Adrichemet al., Anal. Chem., 70:923-930, 1998.

In a preferred aspect, for the analysis of digested proteins, a liquid-chromatography tandem mass spectrometer (LC-MS-MS) is used. This system provides an additional stage of sample separation via use of a liquid chromatograph followed by tandem mass spectrometry.

Figure 2 shows a flowchart depicting the steps carried out in a preferred aspect of the
25   invention for fractions obtained from the second reverse phase HPLC column (RP2) as discussed above and in Example 1. In preferred aspects, a protein eluted from a column according to the system of the invention is analyzed using both MS and MS-MS analysis. For example, a small portion of intact proteins eluting from RP2 may be diverted to online detection using LC-ESI MS. The proteins are aliquoted on a number of plates allowing digestion or not with trypsin,
30   preparation for MALDI-MS as well as for ESI-MS, as well as preparation of the MALDI plates with different matrices. The methods thus allow, in addition to information on intact mass, to conduct an analysis by both peptide mass fingerprinting and MS-MS techniques.

The methods described herein of separating and fractionating proteins provide individual proteins or fractions containing small numbers of distinct proteins. These proteins can be

identified by mass spectral determination of the molecular masses of the protein and peptides

resulting from the fragmentation thereof. Making use of available information in protein

sequence databases, a comparison can made between proteolytic peptide mass patterns generated

in silico, and experimentally-observed peptide masses. A "hit-list" can be compiled, ranking

5   candidate proteins in the database, based on (among other criteria) the number of matches

between the theoretical and experimental proteolytic fragments. Several Web sites are accessible

that provide software for protein identification on-line, based on peptide mapping and sequence

database search strategies (e.g., http://www.expasy.ch). Methods of peptide mapping and

sequencing using MS are described in WO 95/252819, U.S. Pat. No. 5,538,897, U.S. Pat. No.

10   5,869,240, U.S. Pat. No. 5,572,259, and U.S. Pat. No. 5,696,376. See, also, Yates, J. Mass Spec.,

33:1 (1998).

         Data collected from a mass spectrometer typically comprises the intensity and mass to

charge ratio for each detected event. Spectral data can be recorded in any suitable form,

including, e.g., graphical, numerical, or electronic formats, either in digital or analog form.

15   Spectra are preferably recorded in a storage medium, including, e.g., magnetic, such as floppy

disk, tape, or hard disk; optical, such as CD-ROM or laser-disc; or, ROM-CHIPS.

         The mass spectrum of a given sample typically provides information on protein intensity,

mass to charge ratio, and molecular weight. In preferred embodiments of the invention, the

molecular weights of proteins in the sample are used as a matching criterion to query a database.

20   The molecular weights are calculated conventionally, e.g., by subtracting the mass of the ionizing

proton for singly-charged protonated molecular ions, by multiplying the measured mass-over-

charge-ratio by the number of charges for multiply-charged ions and subtracting the number of

ionizing protons.

         Various databases are useful in accordance with the present invention. Useful databases

25   include databases containing genomic sequences, expressed gene sequences, and/or expressed

protein sequences. Preferred databases contain nucleotide sequence-derived molecular masses of

proteins present in a known organism, organ, tissue, or cell-type. There are a number of

algorithms to identify open reading frames (ORF) and convert nucleotide sequences into protein

sequence and molecular weight information. Several publicly accessible databases are available,

30   including, the SwissPROT/TrEMBL database (http://www.expasy.ch).

         Typically, a mass spectrometer is equipped with commercial software that identifies

peaks above a certain threshold level, calculates mass, charge, and intensity of detected ions.

Correlating molecular weight with a given output peak can be accomplished directly from the

spectral data, i.e., where the charge on an ion is one and the molecular weight is therefore equal to

the numerator value minus the mass of the ionizing proton. However, protein ions can be complexed with various counter-ions and adducts, such as N, C, and K'. In such a case, it would be expected that a given protein ion would exhibit multiple peaks, such as a triplet, representing different ionic states (or species) of the same protein. Thus, it may be necessary to analyze and

5   process spectral data to determine families of peaks arising from the same protein. This analysis can be carried out conventionally, e.g., as described by Mann et al., anal. Chem., 61:1702-1708 (1989).

In matching a molecular mass calculated from a mass spectrometer to a molecular mass predicted from a database, such as a genomic or expressed gene database, post-translation

10   processing may have to be considered. There are various processing events which modify protein structure, including, proteolytic processing, removal of N-terminal methionine, acetylation, methylation, glycosylation, phosphorylation, etc.

A database can be queried for a range of proteins which match the molecular mass of the unknown. The range window can be determined by the accuracy of the instrument, the method by

15   which the sample was prepared, etc. Based on the number of hits (where a hit is match) in the spectrum, the unknown protein or peptide is identified or classified.

2D Gel Electrophoresis

In preferred embodiments, the chromatographic separations described herein results in

20   protein fractions (116) from which lower molecular weight proteins and peptides (e.g. preferably having molecular weight of less than e.g. 30kD, 20kD or 15kD) are identified.

However, since it would be desirable to identify all the proteins in a proteome, the methods and systems of the invention combine said method for separating lower molecular weight proteins with a method for separating higher molecular weight proteins. Thus, in

25   preferred embodiments, the methods of the invention based on chromatographic separation further comprise separating proteins from a biological sample using two-dimensional electrophoretic means.

The biological fluid sample to be separated using said two dimensional electrophoresis based method may or may not be pretreated so as to enrich in higher molecular weight proteins,

30   since low molecular weight proteins can flow through the electrophoresis gel and be concentrated at one end without significantly impeding separation of higher molecular weight proteins.

Separation of higher molecular weight proteins can be accomplished according to several strategies. In a preferred method, a fluid sample is treated so as to selectively remove at least one

abundant protein as described above, preferably by passage over an affinity chromatography column. Examples of abundant proteins that can be removed include human serum albumin and immunoglobulins.

     A fluid sample is then subjected to 2-dimensional (2D) gel electrophoresis in order to
5   separate proteins for subsequent detection, preferably by MS methods. Preferably, a pooled biological sample is divided into two or more portions by volume as described above (e.g. having substantially equal protein composition). At least one of the portions is subjected to 2D gel electrophoresis, while at least one of the portions is subjected to liquid chromatographic separation as described above. In a preferred embodiment, the portion of biological sample
10   subjected to 2D gel electrophoresis is not first subjected to gel filtration chromatography. In another example, a portion of sample subjected to size exclusion chromatography to remove proteins below 20 kD is subsequently analyzed by 2D gel electrophoresis.

    Methods and systems for separation by 2D gel electrophoresis are well known in the art; an exemplary protocol is provided in International Patent Publication No. WO 01/63293. Protein
15   samples to be analyzed using 2-D electrophoresis are typically solubilized in an aqueous, denaturing solution such as 9M urea, 2% NP-40 (a non-ionic detergent), 2% of a pH 8-10.5 ampholyte mixture and 1% dithiothreitol (DTT). The urea and NP-40 serve to dissociate complexes of proteins with other proteins and with DNA, RNA, etc. The ampholyte mixture serves to establish a high pH (9) outside the range where most proteolytic enzymes are active,
20   thus preventing modification of the sample proteins by such enzymes in the sample, and also complexes with DNA present in the nuclei of sample cells, allowing DNA-binding proteins to be released while preventing the DNA from swelling into a viscous gel that interferes with IEF separation. The purpose of the DTT is to reduce disulfide bonds present in the sample proteins, thus allowing them to be unfolded and assume an open structure optimal for separation by
25   denaturing IEF. Samples of tissues, for example, are solubilized by rapid homogenization in the solubilizing solution, after which the sample is centrifuged to pellet insoluble material and DNA, and the supernatant collected for application to the IEF gel.

    Because of the likelihood that protein cysteine residues will be come oxidized to cysteic acid or recombine and thus stabilize refolded, not fully denatured protein structures during the
30   run, it is desirable to chemically derivatize the cysteines before analysis. This is typically accomplished by alkylation to yield a less reactive cysteine derivative.

    When used as part of a typical 2-D procedure, an IEF gel is applied along one exposed edge of such a slab gel and the proteins it contains migrate into the gel under the influence of an applied electric field. The IEF gel may be equilibrated with solutions containing SDS, buffer and

thiol reducing agents prior to placement on the SDS gel, in order to ensure that the proteins the

IEF gel contains are prepared to begin migrating under optimal conditions, or else this

equilibration may be performed *in situ* by surrounding the gel with a solution or gel containing

these components after it has been placed in position along the slab's edge. A slab gel affixed to a

5   Gelbond sheet is typically run in a horizontal position, lying flat on a horizontal cooling plate

with the Gelbond sheet down and the unbonded surface up. Electrode wicks communicating with

- liquid buffer reservoirs, or bars of buffer-containing gel, are placed on opposite edges of the slab

to make electrical connections for the run, and samples are generally applied onto the top surface

of the slab (as in the instructions for the Pharmacia ExcelGels).

10.   It is current practice to detect proteins in 2-D gels either by staining the gels or by

exposing the gels to a radiosensitive film or plate (in the case of radioactively labeled proteins).

           Staining methods include dye-binding (e.g., Coomassie Brilliant Blue), silver stains (in

which silver grains are formed in protein-containing zones), negative stains in which, for

example, SDS is precipitated by Zn ions in regions where protein is absent, or the proteins may

15   be fluorescently labeled. In each case, images of separated protein spot patterns can be acquired

by scanners, and this data reduced to provide positional and quantitative information on sample

protein composition through the action of suitable computer software.

20   **EXAMPLES**

Example 1: Separation Protocol for a starting volume of 2500 ml

A total volume of 2.5 liters of plasma was treated using a protocol as follows:

25   **Step 1: HSA/IgG depletion**

           125 ml frozen plasma were defrost and filtered on 0.45 μm sterile filter in a sterile hood.

           Filtrate was injected on two inline columns of respectively 300 ml of HSA ligand

Sepharose fast Flow column (Amersham, Upsala, Sweden), 5cm ID, 15 cm length; and 100 ml

Protein G Sepharose fast Flow column (Amersham, Upsala, Sweden), 5 cm ID, 5 cm length.

30   Columns were equilibrated and washed with 50 mM PO4 buffer, pH 7.1, 0.15M NaCl.

Flow rate was 5 ml/min.

           Non-retained fraction (350 ml) was frozen until second step. 20 runs were performed.

Step 2: Gel Filtration / Reverse Phase Capture step

Sample from step 1 was defrost and filtered on 0.45 μm sterile filter in a sterile hood.

Filtrate was injected on two in line gel filtration columns: 2 X 9.5 litres Superdex 75

(Amersham, UK) column, 14 cm ID, 62 cm length. Column was equilibrated with 50mM PO4

5   buffer pH 7.4, 0.1 M NaCl, 8M urea. Hydrophobic impurities were retained on a reverse phase

precolumn: 150 ml PLRPS (Polymer Labs, UK). Precolumn was switched for sample injection.

Gel filtration was performed at a flow rate of 40 ml/min.

Low molecular weight proteins (<20 kDa) were oriented to in line reverse phase capture

column: 50 ml PLRPS 100 angstroms (Polymer labs, UK). The three-way valve controlling

10   injection on PLRPS column was switched at a cut-off of 33 mAU (280 nm) to send gel filtration

eluate into reverse phase capture column. This cut-off value was established first by SDS-PAGE

to provide an estimated range of OD values and then, by evaluating three cut-off values (high,

median and low values of OD range). The cut-off value was chosen to maximize the low

molecular weight protein obtained, with a low molecular protein proportion of at least 85%. Low

15   molecular weight proteins and peptides were eluted from reverse phase capture PLRPS column

by one column volume gradient of 0.1% TFA, 80% CH3CN in water.

Eluate fractions (50 ml) were frozen until next step. 20 runs were performed. At the end

of this step, all reverse phase eluates were defrost, pooled (1 liter) and shared in 7 polypropylene

containers (143 ml). Containers were kept at –20°C until use for next step.

20

Step 3: Cation Exchange

Sample from step 2 (147 ml) was defrost and mixed with an equal volume of cation

exchange buffer A (Gly/HCl buffer 50 mM, pH 2.7, urea 8M).

Sample was injected on a 100 ml Source 15S column (Amersham, Upsala, Sweden), 35

25   mm ID, 100 mm length. Column was equilibrated and washed with buffer A. Flow rate was 10

ml/min.

Proteins and peptides were eluted with step gradient from 100% buffer A until 100 %

buffer B (buffer A containing 1M NaCl):

3 column volumes 7.5% B (75 mM NaCl)

30   3 column volumes 10% B (100 mM NaCl)

3 column volumes 17.5% B (175 mM NaCl)

2 column volumes 22.5% B (225 mM NaCl)

2 column volumes 27.5% B (275 mM NaCl)

2 column volumes 100% B (1 M NaCl)

45 to 60 fractions were collected based on peak. 7 runs were realized. After 7 runs were achieved, fractions were pooled intra and inter run in order to obtain 18 fractions. Fractions were kept at –20°C until use for next step.

5

Step 4: Reduction/Alkylation and Reverse Phase HPLC Fractionation 1

After adjusting the pH to 8.5 with concentrated tris-HCl, each 18 cation exchange fractions was reduced with dithioerythritol (DTE, 30 mM, 3 hours at 37°C) and alkylated with iodoacetamid (120 mM, 1 hour 25°C in the dark). The latter reaction was stopped with the

10   addition of DTE (30 mM) followed by acidification (TFA, 0.1%). The fractions were then injected on an Uptispher C8, 5 μm, 300 angstroms column (Interchim, France), 21 mm ID, 150 mm length. Injection was performed with a 10 ml/min flow rate.

C8 column was equilibrated and washed with 0.1 TFA in water (solution A). Proteins and peptides were eluted with a biphasic gradient from 100% A until 100% B (0.1% TFA, 80%

15   CH3CN in water) in 60 min. Flow rate was 20 ml/min. 30 fractions of 40 ml were collected.

Based on the measured optical density (OD) at 215 nm of each fraction, which reflects the protein concentration in that fraction, aliquots of similar protein content are created for each fraction.

All aliquots are frozen and kept for further use except one per fraction which is dried with

20   a Speed Vac (Savant, Fischer, Geneva) after addition of 500 μl 10% glycerol in water in each fraction, in order to prevent excess drying. Dried fractions were kept at –20°C until use for next step.

Step 5: Reverse Phase HPLC Fractionation 2

25   Dried samples from step 4 were resuspended in 1 ml of solution A (0.03% TFA in water) and injected on a Vydac LCMS C4 column, 5 micrometers, 300 angstroms (Vydac, USA), 4.6 mm ID, 150 mm length. Flow rate was 0.8 ml/min.

C4 column was equilibrated and washed with solution A and proteins and peptides were eluted with a biphasic gradient adapted to elution position of the sample in Reverse Phase HPLC

30   Fractionation 1. Intact mass data were acquired using Electrospray Ion Trap Mass spectrometry. 16 different gradients were used with a CH3CN concentration range minus and plus 5% CH3CN of RP1 fraction corresponding solvent concentration. For proteins eluted in RP1 with a solvent concentration equal to or greater than 30% CH3CN, the starting elution conditions for the RP2 gradient were set, in CH3CN precentage, at the RP1 elution concentration minus 30%. 24 eluted

fractions were collected in a deep well plate, adopting specific collection configurations, based on CH3CN content, designed for optimal SpeedVac concentration. The evaporation times are also adjusted according to the CH3CN content, and the pressure during the run is monitored to better determine the end of the run.

5

Step 6: Mass detection

Mass detection following the second reverse phase HPLC is carried out as shown in FIG 2.

About 13,000 fractions were collected following reverse phase HPLC fractionation 2 into

10   96-well deep well plates (DWP). A small proportion (2.5 %) of the volume is diverted to online analysis using LC-ESI-MS (Bruker Esquire). Aliquots of undigested proteins are mixed with MALDI matrices, and spotted on MALDI plates together with mass calibration standards and sensitivity standards. Automated spotting devices (Bruker MALDI sample prep. Robots) are used. Two different MALDI matrices are employed: sinapic acid (SA), also known as sinapinic

15   acid, trans-3,5-dimethoxy-4-hydroxycinnamic acid, and alpha-cyano-4-hydroxycinnamic acid (HCCA). MALDI plates are subjected to mass detection using Bruker Reflex III MALDI MS apparati. The 96-well plates are stored at +4 C.

96-well plates (DWP) are recovered and subjected to two sequential concentration steps. Volumes are concentrated from 0.8 ml to about 50 microl per well by drying with a SpeedVac,

20   and then resolubilized to *ca.* 200 microl and reconcentrated to about 50 microl per well, and stored at +4 C. Proteins are then digested by re-buffering, adding trypsin to the wells, sealing and incubating the plates at 37 C for 12 hours, followed by quenching. Quenching is accomplished by bringing the pH to 2 with formic acid. The concentration of trypsin to be added to the wells was adjusted based on the OD at 280 nm recorded for each particular fraction. This ensures an optimal

25   use of trypsin and a complete digestion of the most concentrated fractions. Automated spotting devices (Bruker MALDI sample prep. Robots) are used to deposit a volume from each well, pre-mixed with a HCCA matrix onto a MALDI plate together with sensitivity and mass calibration standards. MALDI plates are analyzed using a Bruker Reflex III MALDI MS device. Contents from each well of the 96 well plates are analyzed LC-ESI-MS-MS Bruker Esquire ESI Ion-Trap

30   MS devices.

Example 2: Detection of low abundance proteins in human plasma

The separation and detection of proteins in human plasma was carried using the process of steps 1 to 6 from Example 1. Intact mass data, Peptide Mass Fingerprints and peptide sequence data were integrated for protein identification and characterization.

5   Proteins were identified using Mascot software (Matrix Science Ltd., London, UK), and results from peptide identification were checked by manual analysis of the spectra.

Exemplary peptides identified included bradykinin, ghrelin, IBP3, Leptin, SY14, SY15 (Pardigol et al.; Proc. Natl. Acad. Sci. USA, 95 (1998) 6308-6313), and SY16 (Hedrick et al.; Blood, 91(1998) 4242-4247), each of which represent relatively low-abundance proteins. SY14

10   (Schulz-Knappe et al., J. Exp. Med., 183 (1996) 295-299) has been reported to have a blood concentration of 1 to 80 nM. IBP3 (Zapf et al.; J. Biol. Chem., 265 (1990) 14892-14898), has been reported to have a blood concentration of 1 nM. Leptin (Zhang et al.; Nature, 372 (1994) 425-432) is known to have a blood concentration of 100 pM. Bradykinin (Blais et al.; Peptides, 21(12) 2000, 1903-1940) of 50 to 500 pM, and ghrelin (Kojima et al.; Nature, 402 (1999) 656-

15   660), of 220 pM.

As an exemplary illustration, Figures 3 to 5 highlight the identification and characterization of leptin from human plasma sample by the method of the invention. Figure 3 shows the MALDI-MS peptide mass for Leptin and Figure 4 shows results from the MS/MS characterization. Finally, Figure 5 displays a summary of the portions of the sequence of leptin

20   that were characterized by mass spectrometry in this example.

Example 3: Separation Protocol for a starting volume of 500 ml

A total volume of 0.5 liter of plasma was treated using a protocol as follows:

25

Step 1: HSA/IgG depletion

125 ml frozen plasma were defrost and filtered on 0.45 µm sterile filter in a sterile hood.

Filtrate was injected on two inline columns of respectively 300 ml of HSA ligand Sepharose fast Flow column (Amersham, Upsala, Sweden), 5cm ID, 15 cm length; and 100 ml

30   Protein G Sepharose fast Flow column (Amersham, Upsala, Sweden), 5 cm ID, 5 cm length.

Columns were equilibrated and washed with 50 mM PO4 buffer, pH 7.1, 0.15M NaCl. Flow rate was 5 ml/min.

Non-retained fraction (350 ml) was frozen until second step. 4 runs were performed.

Step 2: Gel Filtration / Reverse Phase Capture step

Sample from step 1 was defrost and filtered on 0.45 μm sterile filter in a sterile hood.

Filtrate was injected on two in line gel filtration columns: 2 X 9.5 litres Superdex 75 (Amersham, UK) column, 14 cm ID, 62 cm length. Column was equilibrated with 50mM PO4

5   buffer pH 7.4, 0.1 M NaCl, 8M urea. Hydrophobic impurities were retained on a reverse phase precolumn: 150 ml PLRPS (Polymer Labs, UK). Precolumn was switched for sample injection. Gel filtration was performed at a flow rate of 40 ml/min.

Low molecular weight proteins (<20 kDa) were oriented to in line reverse phase capture column: 50 ml PLRPS 100 angstroms (Polymer labs, UK). The three-way valve controlling

10   injection on PLRPS column was switched at a cut-off of 33 mAU (280 nm) to send gel filtration eluate into reverse phase capture column. This cut-off value was established first by SDS-PAGE to provide an estimated range of OD values and then, by evaluating three cut-off values (high, median and low values of OD range). The cut-off value was chosen to maximize the low molecular weight protein obtained, with a low molecular protein proportion of at least 85%. Low

15   molecular weight proteins and peptides were eluted from reverse phase capture PLRPS column by one column volume gradient of 0.1% TFA, 80% CH3CN in water.

Eluate fractions (50 ml) were frozen until next step. 4 runs were performed. At the end of this step, all reverse phase eluates were defrost, pooled (200 ml) and shared in 2 polypropylene containers. Containers were kept at –20°C until use for next step.

20

Step 3: Cation Exchange

Sample from step 2 (100 ml) was defrost and mixed with two volumes of cation exchange buffer A (Gly/HCl buffer 50 mM, pH 2.7, urea 8M).

Sample was injected on a 100 ml Source 15S column (Amersham, Upsala, Sweden), 35

25   mm ID, 100 mm length. Column was equilibrated and washed with buffer A. Flow rate was 10 ml/min.

Proteins and peptides were eluted with step gradient from 100% buffer A until 100 % buffer B (buffer A containing 1M NaCl):

3 column volumes 7.5% B (75 mM NaCl)

30   3 column volumes 10% B (100 mM NaCl)

3 column volumes 17.5% B (175 mM NaCl)

2 column volumes 22.5% B (225 mM NaCl)

2 column volumes 27.5% B (275 mM NaCl)

2 column volumes 100% B (1 M NaCl)

45 to 60 fractions were collected based on peak. 2 runs were realized. After 2 runs were achieved, fractions were pooled intra and inter run in order to obtain 12 fractions. Fractions were kept at –20°C until use for next step.

5

Step 4: Reduction/Alkylation and Reverse Phase HPLC Fractionation 1

After adjusting the pH to 8.5 with concentrated tris-HCl, each 12 cation exchange fractions was reduced with dithioerythritol (DTE, 30 mM, 3 hours at 37°C) and alkylated with iodoacetamid (120 mM, 0.5 hour 37°C in the dark). The latter reaction was stopped with the

10   addition of DTE (30 mM) followed by acidification (TFA, 0.1%). The fractions were then injected on an Uptispher C8, 5 µm, 300 angstroms column (Interchim, France), 21 mm ID, 150 mm length. Injection was performed with a 10 ml/min flow rate.

C8 column was equilibrated and washed with 0.1 TFA in water (solution A). Proteins and peptides were eluted with a biphasic gradient from 100% A until 100% B (0.1% TFA, 80%

15   CH3CN in water) in 30 min. Flow rate was 20 ml/min. 15 fractions of 40 ml were collected.

Based on the measured optical density (OD) at 215 nm of each fraction, which reflects the protein concentration in that fraction, aliquots of similar protein content are created for each fraction.

All aliquots are frozen and kept for further use except one per fraction which is dried with

20   a Speed Vac (Savant, Fischer, Geneva) after addition of 500 µl 10% glycerol in water in each fraction, in order to prevent excess drying. Dried fractions were kept at –20°C until use for next step.

Step 5: Reverse Phase HPLC Fractionation 2

25   Dried samples from step 4 were resuspended in 1 ml of solution A (0.03% TFA in water) and injected on a Vydac LCMS C4 column, 3 micrometers, 300 angstroms (Vydac, USA), 4.6 mm ID, 100 mm length. Flow rate was 0.8 ml/min.

C4 column was equilibrated and washed with solution A and proteins and peptides were eluted with a biphasic gradient adapted to elution position of the sample in Reverse Phase HPLC

30   Fractionation 1. Intact mass data were acquired using Electrospray Ion Trap Mass spectrometry. 16 different gradients were used with a CH3CN concentration range minus and plus 5% CH3CN of RP1 fraction corresponding solvent concentration. For proteins eluted in RP1 with a solvent concentration equal to or greater than 30% CH3CN, the starting elution conditions for the RP2 gradient were set, in CH3CN precentage, at the RP1 elution concentration minus 3%. 24 eluted

fractions were collected in a deep well plate, adopting specific collection configurations, based on CH3CN content, designed for optimal SpeedVac concentration. The evaporation times are also adjusted according to the CH3CN content, and the pressure during the run is monitored to better determine the end of the run.

5

Step 6: Mass detection

4320 fractions were collected following reverse phase HPLC fractionation 2 into 96-well deep well plates (DWP). A small proportion (2.5 %) of the volume is diverted to online analysis using LC-ESI-MS (Bruker Esquire). Aliquots of undigested proteins are mixed with MALDI

10   matrices, and spotted on MALDI plates together with mass calibration standards and sensitivity standards. Automated spotting devices (Bruker MALDI sample prep. Robots) are used. Two different MALDI matrices are employed: sinapic acid (SA), also known as sinapinic acid, trans-3,5-dimethoxy-4-hydroxycinnamic acid, and alpha-cyano-4-hydroxycinnamic acid (HCCA). MALDI plates are subjected to mass detection using Bruker Reflex III MALDI MS apparati. The

15   96-well plates are stored at +4 C.

96-well plates (DWP) are recovered and subjected to two sequential concentration steps. Volumes are concentrated from 0.8 ml to about 50 microl per well by drying with a SpeedVac, and then resolubilized to *ca.* 200 microl and reconcentrated to about 50 microl per well, and stored at +4 C. Proteins are then digested by re-buffering, adding trypsin to the wells, sealing and

20   incubating the plates at 37 C for 12 hours, followed by quenching. Quenching is accomplished by bringing the pH to 2 with formic acid. The concentration of trypsin to be added to the wells was adjusted based on the OD at 280 nm recorded for each particular fraction. This ensures an optimal use of trypsin and a complete digestion of the most concentrated fractions. Automated spotting devices (Bruker MALDI sample prep. Robots) are used to deposit a volume from each well, pre-

25   mixed with a HCCA matrix onto a MALDI plate together with sensitivity and mass calibration standards. MALDI plates are analyzed using a Bruker Reflex III MALDI MS device. Contents from each well of the 96 well plates are analyzed LC-ESI-MS-MS Bruker Esquire 3000 plus ESI Ion-Trap MS devices.

30

Example 4: Separation Protocol for a starting volume of 10 ml
A total volume of 10 ml of plasma was treated using a protocol as follows:

Step 1: HSA/IgG depletion

      10 ml frozen plasma were defrost and filtered on 0.45 μm sterile filter in a sterile hood.

      Filtrate was injected on two inline columns of respectively 30 ml of HSA ligand

Sepharose fast Flow column (Amersham, Upsala, Sweden), 1.6 cm ID, 15 cm length; and 10 ml

5   Protein G Sepharose fast Flow column (Amersham, Upsala, Sweden), 1.6 cm ID, 5 cm length.

      Columns were equilibrated and washed with 50 mM PO4 buffer, pH 7.1, 0.15M NaCl.

Flow rate was 0.5 ml/min.

      Non-retained fraction (35 ml) was frozen until second step.

10   **Step 2: Gel Filtration / Reverse Phase Capture step**

      Sample from step 1 was defrost and filtered on 0.45 μm sterile filter in a sterile hood.

      Filtrate was injected on three in line gel filtration columns: 3 X 0.6 litres Superdex 75

(Amersham, UK) column, 4.4 cm ID, 40 cm length. Columns were equilibrated with 50mM PO4

buffer pH 7.4, 0.1 M NaCl, 8M urea. Hydrophobic impurities were retained on a reverse phase

15   precolumn: 15 ml PLRPS (Polymer Labs, UK). Precolumn was switched for sample injection.

Gel filtration was performed at a flow rate of 4 ml/min.

      Low molecular weight proteins (<20 kDa) were oriented to in line reverse phase capture

column: 5 ml PLRPS 100 angstroms (Polymer labs, UK). The three-way valve controlling

injection on PLRPS column was switched at a cut-off of 33 mAU (280 nm) to send gel filtration

20   eluate into reverse phase capture column. This cut-off value was established first by SDS-PAGE

to provide an estimated range of OD values and then, by evaluating three cut-off values (high,

median and low values of OD range). The cut-off value was chosen to maximize the low

molecular weight protein obtained, with a low molecular protein proportion of at least 85%. Low

molecular weight proteins and peptides were eluted from reverse phase capture PLRPS column

25   by one column volume gradient of 0.1% TFA, 80% CH3CN in water.

      Eluate fraction (5 ml) was frozen until next step.

Step 3: Cation Exchange

      Sample from step 2 (5 ml) was defrost and mixed with two volumes of cation exchange

30   buffer A (Gly/HCl buffer 50 mM; pH 2.7, urea 8M).

      Sample was injected on a 10 ml Source 15S column (Amersham, Upsala, Sweden), 1 cm

ID, 10 cm length. Column was equilibrated and washed with buffer A. Flow rate was 1 ml/min.

      Proteins and peptides were eluted with step gradient from 100% buffer A until 100 %

buffer B (buffer A containing 1M NaCl):

3 column volumes 7.5% B (75 mM NaCl)

3 column volumes 10% B (100 mM NaCl)

3 column volumes 17.5% B (175 mM NaCl)

2 column volumes 22.5% B (225 mM NaCl)

5   2 column volumes 27.5% B (275 mM NaCl)

2 column volumes 100% B (1 M NaCl)

15 fractions were collected based on time. Fractions were kept at –20°C until use for next

step.

10

Step 4: Reduction/Alkylation and Reverse Phase HPLC Fractionation

After adjusting the pH to 8.5 with concentrated tris-HCl, each 15 cation exchange

fractions was reduced with dithioerythritol (DTE, 30 mM, 2 hours at 37°C) and alkylated with

iodoacetamid (120 mM, 30 min at 37°C in the dark, under agitation). The latter reaction was

15   stopped with the addition of DTE (30 mM) followed by acidification (TFA, 0.1%). The fractions

were then injected on a Vydac C4, 3 μm, 300 angstroms column (Vydac, CA, USA), 4.6 mm ID,

and 100 mm length. C4 column was equilibrated and washed with 0.05 % TFA in water (solution

A). Proteins and peptides were eluted with a biphasic gradient from 100% A until 100% B (0.1%

TFA, 80% CH3CN in water) in 15 min. Flow rate was 0.8 ml/min. 15 fractions of 0.6 ml were

20   collected.

Step 5: Mass detection

225 fractions were collected following reverse phase HPLC fractionation into 96-well

deep well plates (DWP). A small proportion (2.5 %) of the volume is diverted to online analysis

25   using LC-ESI-MS (Bruker Esquire). 96-well plates (DWP) are recovered and subjected to

concentration step. Volumes are concentrated from 0.8 ml to about 50 microl per well by drying

with a SpeedVac and proteins are then digested by re-buffering, adding trypsin to the wells,

sealing and incubating the plates at 37 C for 12 hours. The concentration of trypsin to be added

to the wells was adjusted based on the OD at 210 nm recorded for each particular fraction. This

30   ensures an optimal use of trypsin and a complete digestion of the most concentrated fractions.

Contents from each well of the 96 well plates are analyzed LC-ESI-MS-MS Bruker Esquire ESI

Ion-Trap MS devices.

The descriptions of the foregoing embodiments of the invention have been presented for

5   purpose of illustration and description. They are not intended to be exhaustive or to limit the

invention to the precise forms disclosed, and obviously many modifications and variations are

possible in light of the above teaching. The embodiments were chosen and described in order to

best explain the principles of the invention to thereby enable others skilled in the art to best utilize

the invention in various embodiments and with various modifications as are suited to the

10   particular use contemplated. It is intended that the scope of the invention be defined by the

claims appended hereto. Furthermore, several references have been cited in the present

disclosure. Each of the cited references is incorporated herein by reference.